Optimisation of Automatic Train Protection Systems

Daniel Woodland

A Thesis Submitted for the Degree of Doctor of Philosophy

August 2004

Department of Mechanical Engineering
University of Sheffield

EXECUTIVE SUMMARY

In Britain, Automatic Train Protection (ATP) systems are perceived as expensive and as detrimental to railway capacity. Optimal implementation is thus required in order to be of benefit to railways through improvements in both safety and efficiency.

The author of this thesis reports on an extensive literature review into the background of ATP systems, along with research into methods for optimising their future application. The literature review includes:

- An investigation into the purposes and context of train control;
- An overview of the historical development of train control and of approaches to eliminating human error from the control of train movements;
- Lessons to be learnt from the different systems implemented.

Based on this review, the author identified railway capacity as the most significant area for further research in the context of ATP systems, with cost and safety as secondary issues, to be considered when evaluating options for improving capacity.

Accordingly, detailed research is presented on:

- Train braking performance;
- Speed supervision criteria and
- The capacity impact of implementing ATP systems using different data transmission methods and train separation strategies.

The research results reveal that existing technology can be utilised to enhance steady state capacities on main line railways safely by up to:

- 15% through implementing fixed block in-cab ATP systems;
- 34% through implementing moving block train separation;
- 45% through improved rail / wheel interface management and train braking performance, in conjunction with moving block train separation.

On metro railways, the benefits would be up to:

- 5.6% through implementing moving block train separation;
- 16% through improved rail / wheel interface management and train braking performance, in conjunction with moving block train separation.

In both cases, the full steady state capacity benefit of applying moving block train separation may be achieved by localised application in the vicinity of in-line station stops only.

A fuller summary of the results obtained can be found in Figure S1.

In contrast to these opportunities for capacity enhancement, the author's modelling revealed that overlaying ATP systems on conventional lineside signalling would have significant detrimental effects on achievable railway capacity. Such overlay systems would, therefore, only be appropriate on a line that is lightly used when compared with the design capacity of its existing lineside signals.

The results for both main line and metro railway models demonstrate a good case for the implementation of ATP in a continuous in-cab form and for the adoption of moving block

rather than fixed block train separations, at least in the area of in-line station platforms. Whilst the benefit to be gained from moving block application is higher on main lines than it is on metros, the demand for increased metro capacity on many of the worlds metro railways is sufficient to make the smaller benefit remain highly attractive – particularly since the viability of moving block has been proved by several equipment manufacturers on metro railways throughout the world.

By contrast, modelling of the effects of relative braking separations (that is, where following trains maintain a safe braking distance to the projected stop point of the train ahead, rather than its actual location) revealed that:

- Capacity increase of up to 3% for main line traffic and 5% for metro railways could potentially be gained over those achievable by use of moving block train separations;
- The implementation of relative braking separations would introduce a significant increase in collision risk as a part of implementing unproven technology.

The risk issues associated with relative braking separations would appear to undermine the case for developing moving block into relative braking train separation, despite the attraction of increased capacity.

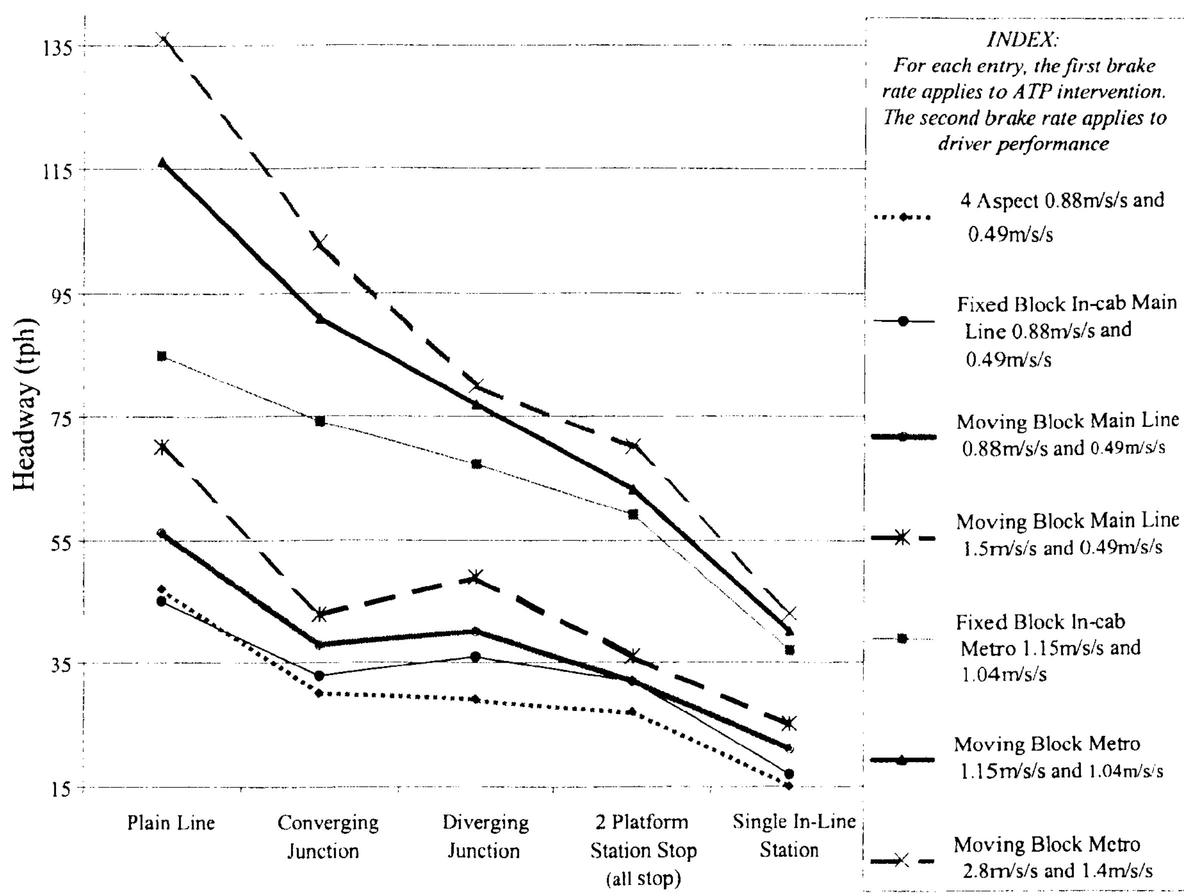


Figure S1: Summary of Potential for Steady State Capacity Enhancement

ACKNOWLEDGEMENTS

I would like to thank:

- My supervisors, Dr Felix Schmid of Sheffield University and Bob Barnard of Alstom Transport Information Solutions, for their guidance and support during the production of this thesis;
- Mariana Somasundaram for encouraging me to undertake the programme of research that it documents and arranging for Alstom Transport Information Solutions to sponsor the first two years of study;
- John Sadler of Bechtel Ltd for taking on the sponsorship of the last two years study;
- My work colleagues in both the West Coast Main Line Joint Project Team and Rail Link Engineering (CTRL) for their support, input and critical review of my thoughts and ideas. In particular, the discussions held with Jerry Lewis were of great assistance to me;
- All of the individuals who allowed me to conduct interviews with them (as documented in Appendix A). I am extremely grateful to each of them (and the companies that they represent).

During the course of the research outlined within this report, I have been overwhelmed by the volume of material made available for the purposes of my research. In particular I would like to mention Paul LeVesconte and Tony Sprought of Rail Training International and Ken Bott of AEA Technology Rail for their assistance in this way.

It is my opinion that a true picture of a system's operation can never be gained from reading about it alone. I am, therefore, extremely grateful to contacts within the rail industry who consented to proof read portions of this report and offer the benefit of advice based on their practical experience of ATP systems. In particular, my gratitude goes to Steve Rodgers of Booze Allen Hamilton, Gilles Poitrison-Rivierre and Tim Brockbank of Alstom Transport Information Solutions and Tom Lee of RSSB.

Finally, I would like to express my thanks to members of my family. I owe my father, Colin Woodland, sincere thanks for writing an Access database for me (which has proved an invaluable tool for recording the results of my research). Most of all, however, my thanks are due to my wife, Julie Woodland, who has encouraged and supported me during the production of this report (putting up with my preoccupation and frequent requests for proof reading!)

CONTENTS

1 INTRODUCTION

In the modern world, freedom of choice is a reality for customers wishing to buy most products. Transport is no exception to this, with the choice of foot, bicycle, boat, aircraft and motor vehicle (whether car, taxi, bus, van or lorry) all being offered in competition with rail for the transportation of both passengers and goods. In the light of this competition, "railways will provide a significant contribution to transport infrastructure only if they deliver a service of the quality demanded by customers at a price which they can afford" (Perry 1998, pM45-1). However, at the current time, both the quality and cost of transport by rail are continually in the media spotlight, invariably receiving unfavourable reviews. This is particularly true when it comes to the subject of Automatic Train Protection (ATP) – a system intended to improve safety, but often considered too costly to apply, both in terms of direct financial cost and of its impact on railway capacity.

The title 'Optimisation of Automatic Train Protection Systems' was chosen by the author because of the perceived need for a way of implementing ATP systems that would benefit Britain's railways, enhancing their competitiveness in a cost effective way through improvements in both safety and efficiency. However, at the start of his study, this was only a vague notion. The author's first task was, therefore, to identify the scope that should be encompassed by the title in order to maximise the benefits that could be gained from his subsequent research.

In order to do this, an extensive literature review was carried out into the background of Automatic Train Protection systems. The boundaries of this review were deliberately set wider than train protection systems alone, in order to provide a flavour of the context within which train protection operates and to assess the scope of further research that would best 'optimise' future train protection systems.

In accordance with this objective, the literature review commenced with an investigation into the purposes of train control systems, considering both the context in which they operate (including that of the railway system in general and of railway control systems in particular) and the competing demands of safety, capacity and economics that are placed upon them. The findings of this investigation are documented in chapter 2 of this thesis.

Once the author had ascertained the purposes of train control systems, he focused next on the historical development of train control and the lessons to be learnt from the different types of systems that have been implemented throughout the history of railway operation. This investigation began with a consideration of 'Signalling' based train control (documented in chapter 3 and Appendix B) and moved on to more advanced methods of train control, such as transmission based signalling and automation of control (documented in chapter 4 and Appendix B).

Having completed the initial review of the context of train protection systems, the author then turned his attention to a more detailed consideration of the approaches adopted historically for eliminating human error from the control of train movements. In order to do this, a variety of manual and automatic train protection systems that have been in use around the world were considered. Descriptions of these systems and the lessons that can be learnt from them are given in Appendix C.

Where published material addressing subjects of interest could not be found during the literature review, the author conducted interviews with experienced professionals in railway signalling, train control, operation and other relevant subject areas. Transcripts of these interviews can be found in Appendix A.

Based on this background research, the author identified railway capacity as being the most significant area for further research, with cost and safety as secondary issues to be considered when evaluating options for improving capacity through optimisation of ATP systems. This led to detailed research into:

- Train braking performance (documented in chapter 5);
- Speed supervision criteria (documented in chapter 6) and;
- The capacity impact of implementing ATP systems with different data transmission methods and train separation strategies (documented in chapters 7 to 11 and Appendices D, E and F).

Throughout the research work, consideration was given to both main line and metro railways, with a predominant focus on British practices.

Summaries of this research have been provided at the end of each chapter and overall conclusions can be found in chapter 12.

Within the time available for completion of this thesis, it has not been possible to thoroughly investigate every area of interest that was identified by the author. Ideas for further research have therefore been outlined in chapter 13.

Full details of referenced sources can be found in the 'References and Bibliography' section, and definitions of abbreviations used in the 'List of Abreviations'. These sections are located after chapter 13.

The author of this thesis has been working in the field of railway signalling and control for 10 years. In addition to the research documented within this thesis, his career has included approximately:

- 5 years working as a signalling application designer, developing scheme plans, control tables and detailed circuit designs (London Underground and Alstom Transport Information Solutions);
- 3 years working as a railway systems engineer, developing principles for the implementation of an ERTMS based ATP system on UK main line railways. This work encompassed issues associated with signalling, communication systems, rolling-stock and railway operations (Alstom Transport Information Solutions);
- 2 years working as a systems integration engineer, developing the system architecture and assisting interface management for all electronic and electrical sub-systems on section 2 of the Channel Tunnel Rail Link (Bechtel Ltd).

2 THE PURPOSE OF TRAIN CONTROL SYSTEMS

2.1 INTRODUCTION

Train control systems exist within the context of an operational railway. Their purpose is to ensure that trains can run safely and efficiently, so as to satisfy the expectations of customers, regulatory authorities and the railway operators.

As the primary purpose of the railway system is concerned with the physical transfer of passengers and goods, it is fundamental to railway control systems that they should be concerned with the positional control of trains (Short 1996, B2/1). Therefore, the essential purpose of any train control system can be summarised as:

a. To maintain a safe distance between following trains on the same track;
b. To safeguard the movement of trains at junctions and where crossing a path which could be taken by another train;
c. To control train movement between and at stations;
d. To regulate the passage of trains according to the service density and speed required, accounting for the planned schedule.

(Khessib 1989, p2; Nock 1980, p1).

This places the train control system at the heart of the railway.

In fulfilling its principal task of ensuring the safe separation of trains, the train control system affects the sort of service that can be operated in terms of capacity and speed (Rowbotham 1999, p1), and also provides a means of regulating the service. The train control system therefore has competing requirements placed upon it: those of safety and those of operational capacity.

A more detailed discussion of the types of system that have historically been used in order to achieve the purposes of train control can be found in chapters 3 and 4. Therefore, further consideration will not be given to the subject within this chapter.

Whereas railway control around the world has traditionally evolved in response to external pressures (such as accidents and technology advances), modern control systems are beginning to consider the system as a whole. According to the Institution of Railway Signal Engineers "it is becoming ever more apparent that an integrated systems approach is needed for the design and operation of a railway, in order that there is clarity about which elements of the railway are responsible for what" (IRSE(1) 2001, paragraph 47). In order to do this in a truly meaningful way, it is necessary to define the system's purpose, how it fits into its environment and how it will treat its customers and staff.

The author therefore begins this chapter by considering the context in which the train control system operates. He then goes on to consider the primary requirements and hazards of railway operation, together with techniques for assessing the feasibility of a proposed system.

2.2 THE RAILWAY SYSTEM

The railway system consists of equipment and people, together with the rules and regulations that govern all activities and operations, whether in a steady state, a degraded mode or another transitional state (IRSE(1) 2001 paragraph 40). The characteristics of rail transport differ from those of other transport modes in a number of ways. The most obvious of these is the inherent restriction of movement to a single degree of freedom - the train being forced by the flanges of its wheels to follow the lateral guidance of a rail. The low coefficient of friction between metal wheels and metal rails, combined with high operating speeds and large mass, also make the braking distance of a train too long for safe operation purely on the basis of a driver's reaction to objects within his/her line of sight. As a result, rail transport can not be operated on the same basis as, say, road traffic – where the driver of an individual vehicle has the sole responsibility for guiding its movement. Instead a control system is required to ensure that movements are carried out in a safe manner (Uebel 1996, pC1/1).

The use of a fixed track infrastructure also makes the number of route options available much smaller than those for road, sea or air transport. This restriction is accentuated by the need for a safe separation to be maintained between trains on the same track segment. As a result, there is a far greater requirement for access planning, or scheduling, to ensure that optimum use is made of the available routes (Barter 2001, p1; Etschmaier 1986, p149).

A more comprehensive overview of the physical characteristics of railways is shown in Table 2-1.

	Motion Restricted to One Degree of Freedom	**Low Coefficient of Friction Between Wheel and Rail**	**Distribution of Load over a Large Area**
Direct Strengths	• No steering required • Predictable motion • Narrow swept path – low land occupation requirements	• Energy efficiency • Smooth operation • Efficiency of propulsion	• Heavy axle loads • High tonnage / period • Low ground forces
Indirect Strengths	• High standard of safety • Use of linked consists (trains) • High capacity • Wayside power supply is possible • Favours use of automatic controls	• High speed operation • Energy recovery • Lower environmental pollution	• High load capacity • Long life infrastructure
Direct Weaknesses	• High initial system costs for guidance and control infrastructure • High infrastructure maintenance costs • Low network flexibility • Can not swerve to avoid a collision • Overtaking and passing movements require physical infrastructure • Limited number of vehicle paths available	• Distance to stop • Distance to accelerate	• Stiff rolling interface • Low inherent damping • Cost of structures • Cost of track
Indirect Weaknesses	• No door to door service • Poor access to shipper's premises • Reliant on performance of connecting modes	• No line of sight working • Seasonal variations in train performance	• Noise generation • Cost of inspection
Requirements	• Junctions and stations • Interlocking of variable geometry infrastructure components • Scheduling of movements in advance • Control systems to ensure efficient operation	• Train control systems to ensure safe operation • Adhesion control • Stringent safety rules	• Track design tailored to loads

Table 2-1: Characteristics of Rail Transport (based on, Schmid [(1)] 2001)

Based on these characteristics, it is the author's opinion that railways require a control system to support their operating schedule, both to maximise the number of train paths that can safely be designed into it and to assist in maintaining a service that reliably achieves the schedule's demands. If such a control system is to be effective in providing this support, it must be designed as an integral part of the railway system.

2.2.1 PROCESS VIEW

The Railway System can be represented in terms of the processes required for its development and operation. This can be done by use of an IDEF0 (Integration Definition for Function Modelling) model.

A top level IDEF0 representation of this process is given in Figure 2-1. Within IDEF0, a process is represented by a rectangular box. Inputs enter this box from the left and outputs leave from the right. Controls / constraints on the process enter from the top of the box and mechanisms providing the means of carrying out the process enter from below.

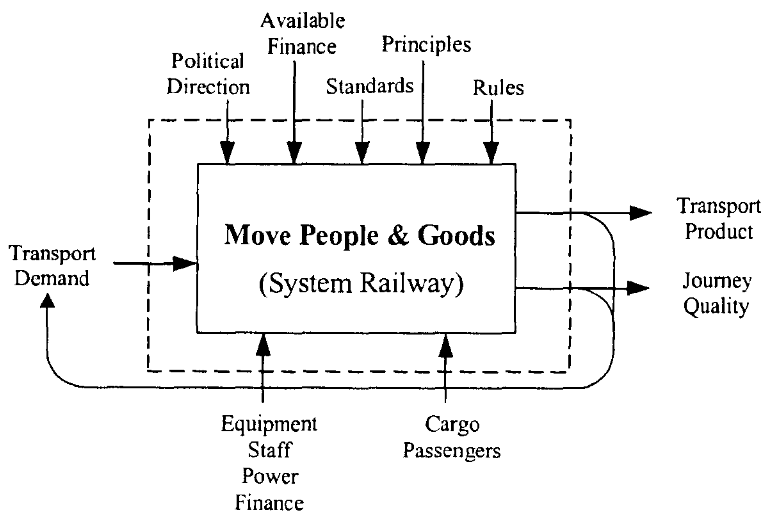


Figure 2-1: Top Level IDEF0 Representation of The Railway System

If this is decomposed to represent the processes that are required to develop and operate a Railway System, the first level IDEF0 representation shown in Figure 2-2 is obtained.

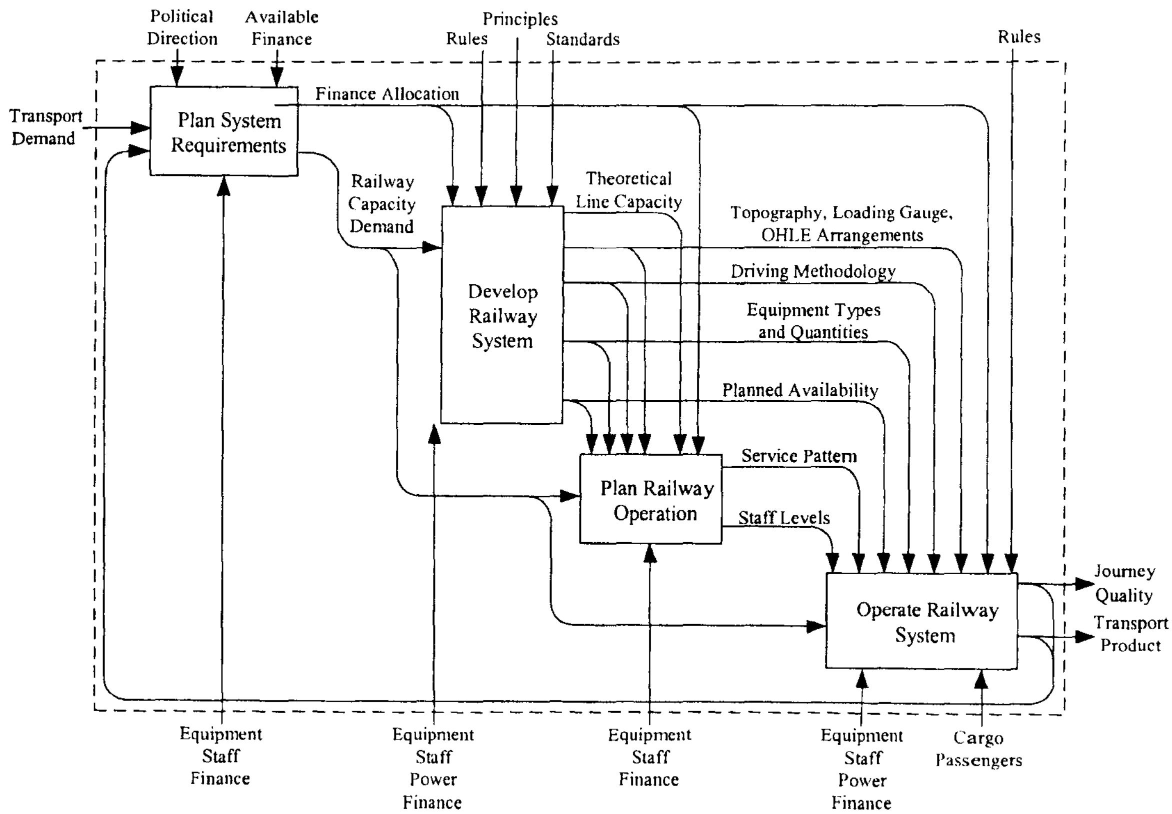


Figure 2-2: First Level View of the Railway System

It should be noted that overall transport demand acts as the initiating input for the entire transportation system, not just the railway system process. There will also be a limit to available finance for transportation as a whole and a given railway system in particular. Some proportion of the transportation demand and available finance must be apportioned to railways and, then, on to specific railway networks. The first process identified is, therefore, the planning of system requirements for the railway. This process must apportion the demand for transportation and the available finance into railway capacity demand and finance allocations. The mechanisms of this process are not of importance to this thesis and will not be expanded further at this stage.

Once the railway capacity demand is known, the railway system itself can be developed in accordance with that demand, as shown in Figure 2-3.

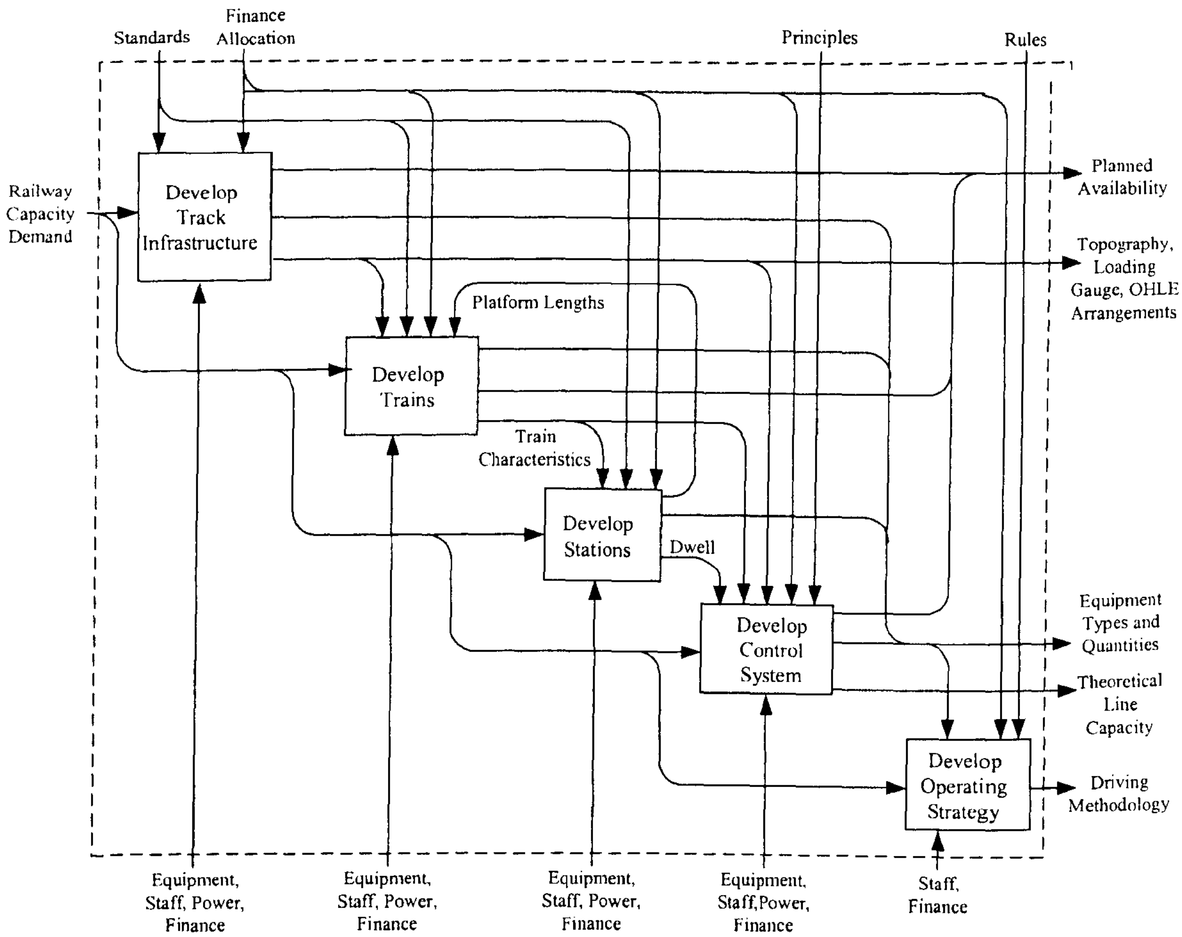


Figure 2-3: Second Level View "Develop Railway System"

Based upon the outcomes of the 'Develop Railway System' process, plans can be drawn up for the railway's operation. These must account for the demands of the systems users, the arrangement and types of both fixed and moveable infrastructure, and staffing issues. The mechanisms of this process are once again not of great importance to this thesis and will not be expanded further. However, once the service pattern and staffing levels have been determined, the actual operation of the railway can be studied, as shown in Figure 2-4.

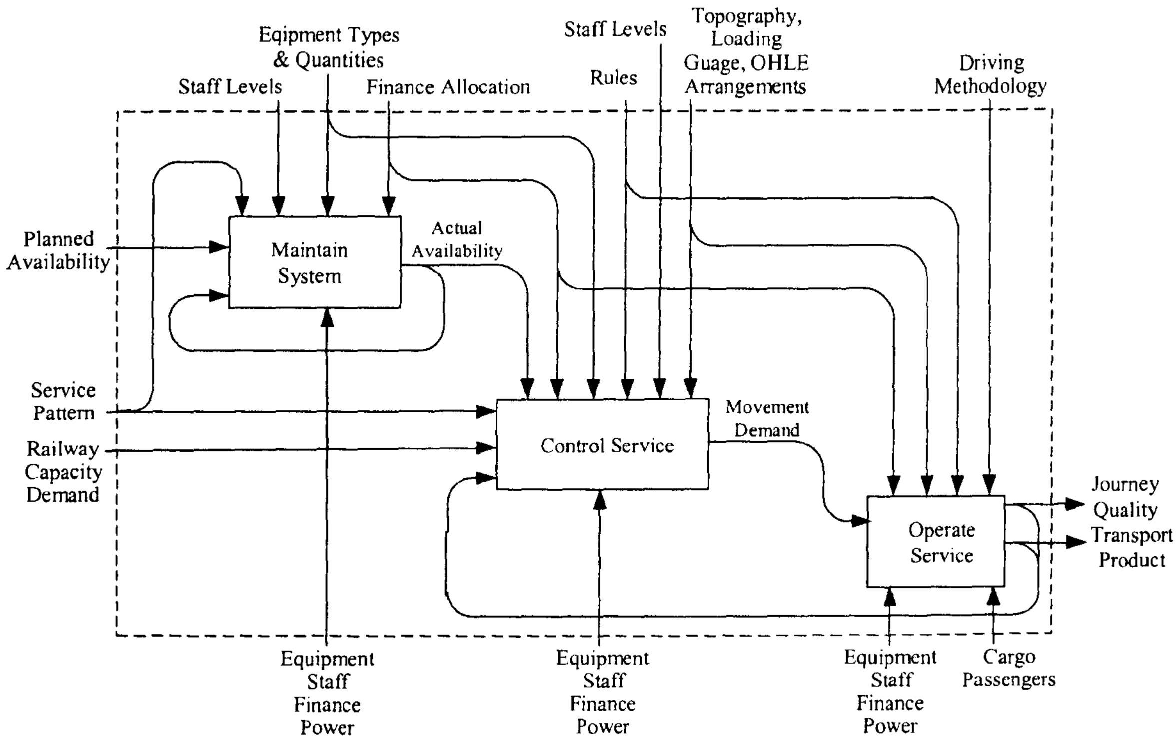


Figure 2-4: Second Level View "Operate Railway System"

The key point to note from these IDEF0 representations of the railway system is that a number of different processes must work together, utilising a number of engineering systems, in order to produce a transport product that meets the customers' demand for transportation in general, and railway capacity in particular. When considering the 'Optimisation of Automatic Train Protection', certain processes (such as 'Develop Railway System' and 'Operate Railway System') may present themselves as having particular relevance. It may even be possible to identify key sub-processes (such as 'Develop Control Systems' and 'Control Service'). However, true 'optimisation' will only be possible if consideration is given to all of the processes that contribute to the production and operation of the railway's control system (and specifically to the Automatic Train Protection system).

2.2.2 SUB-SYSTEM VIEW

If an alternative view is taken, a system can be considered to be formed of a number of interrelated sub-systems, each of which contains a number of interrelated sub-systems of its own. If this method of thinking is applied to a railway system, it can be seen that the 'railway' is itself a sub-system of the transport network that is in turn within the system of a country's economy. In a similar way, the railway system will be composed of sub-systems, including the control system (Goddard 1996, D5/24). The control system cannot operate in isolation to the wider railway system and, therefore, must be able to cope with both planned and unplanned demands that are placed on it by the day-to-day operation of the railway. Figure 2-5 shows a basic outline of this approach, portraying a number of factors, both internal and external to the railway system, that influence (either directly or indirectly) the operation of the control system.

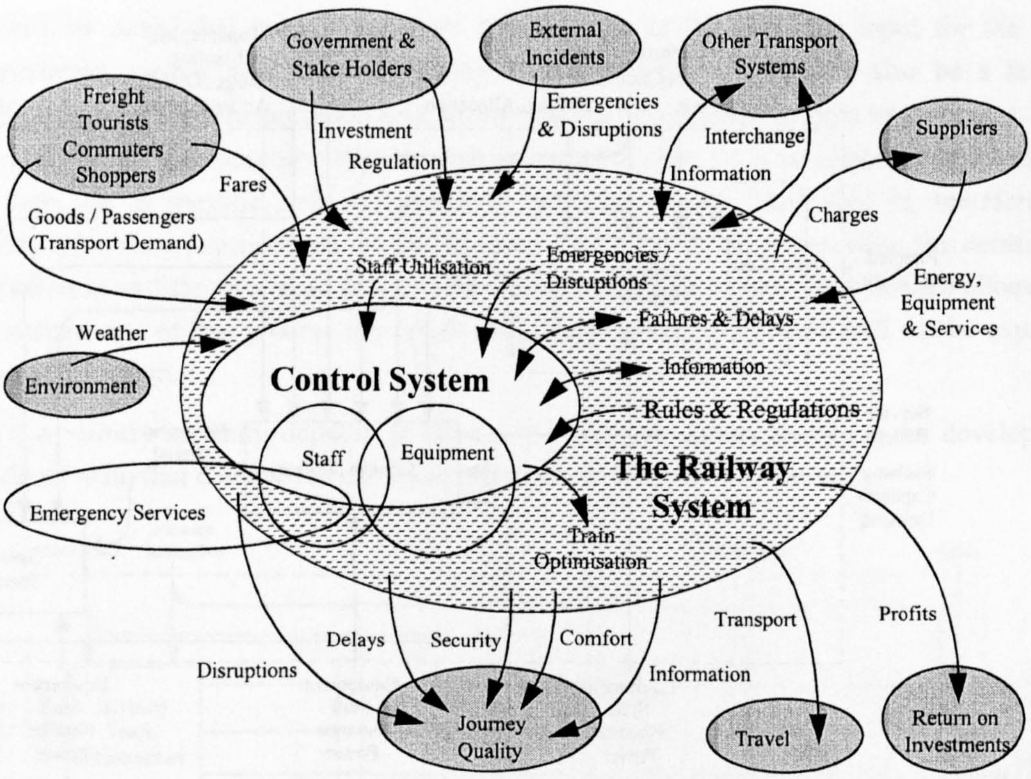


Figure 2-5: The Railway System

2.3 THE RAILWAY CONTROL SYSTEM

The control system of a railway can be viewed in operational terms as a pyramid (Figure 2-6).

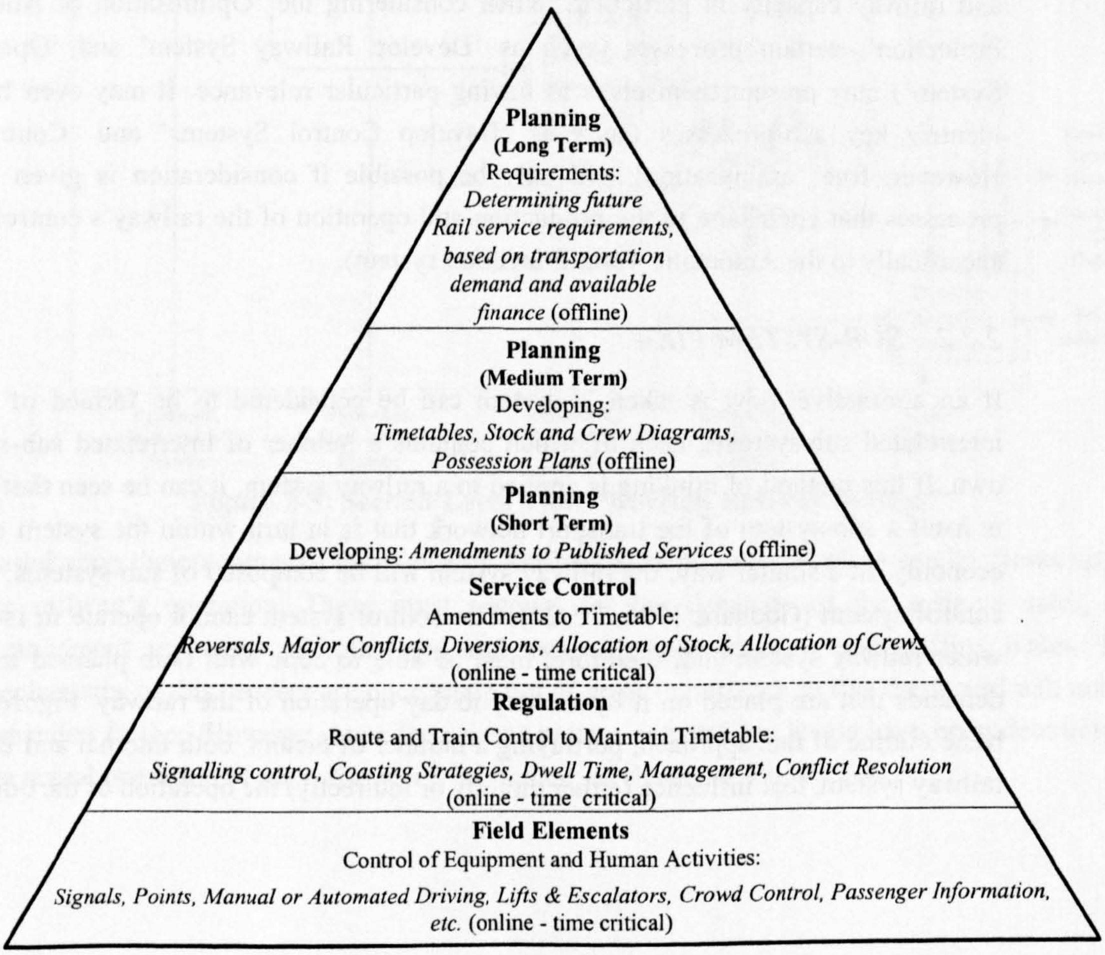


Figure 2-6: Railway Control System – Operational View (based on Barter 2001; Uebel 2000)

The top three elements of the pyramid represent the offline, non-vital, activities. The two sections below these have no strict separation (hence the dotted line) and represent the real-time, vital, elements of the control system. The bottom of the pyramid represents the physical elements of the railway to be controlled. This system model holds true whether the control activities are automated or manual, and whether the field elements are mechanical, electronic or purely human (e.g., a person with a red flag).

The engineering view of the railway control system has not always agreed with the operational view. Engineers (or at least Signalling Engineers) have historically treated railway control as purely concerned with ensuring the safe movement of trains. However, the introduction of modern technology has led to a demand for a more rigorous approach to railway control that incorporates a multitude of control sub-systems related to the comfort and well being of customers (Goddard 1996, D5/16). Hence, the railway control system now needs to include and integrate the control of train movements and stations. It must consider the safe control of operations, but not neglect the operational requirements to keep passengers informed, regulate trains, save energy and provide for optimal usage of the track.

Building on the system outline given in Figure 2-5, the scope of railway control systems can be seen to be more than just the control of train movements. Figure 2-7 provides an outline of the main sub-systems that combine to form the railway control system.

Although two distinct views of railway control, the operational and engineering sub-system views are mutually compatible. Whilst the engineering view shows the types of sub-system that are needed to control the railway successfully, the operational view provides their context and purpose, showing what actually needs to be controlled.

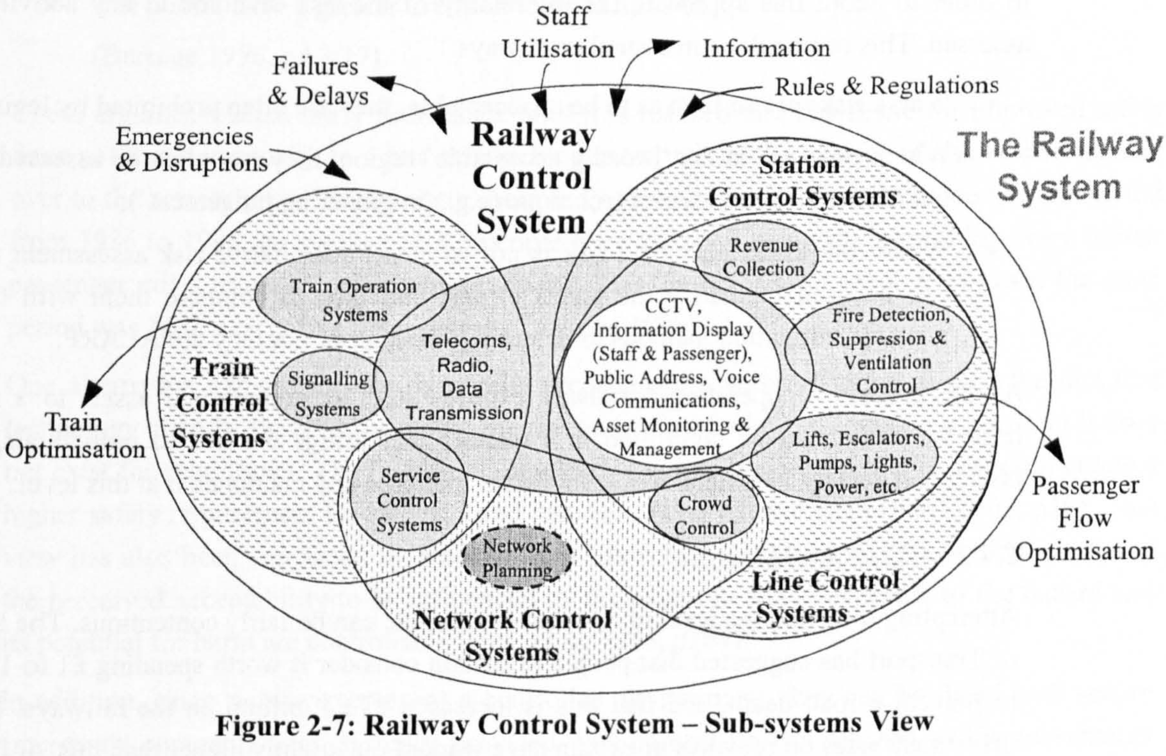


Figure 2-7: Railway Control System – Sub-systems View

2.4 RAILWAY SAFETY

The importance of safe railway operation has long been recognised, as have the practical limitations of economics. In 1910, the inaugural meeting of the Institution of Signal Engineers was advised that the objective of signalling for train control was to “provide for safety without undue sacrifice of economy, whether in time or money" (Blackall 1910, p12). This philosophy is still as relevant today as it was then, and is now commonly referred to as the principle of ALARP. This states, "risks to individuals and society should be As Low As Reasonably Practicable" (The Engineering Council 1993, page 21).

2.4.1 ALARP

The principle of ALARP is explained in Figure 2-8.

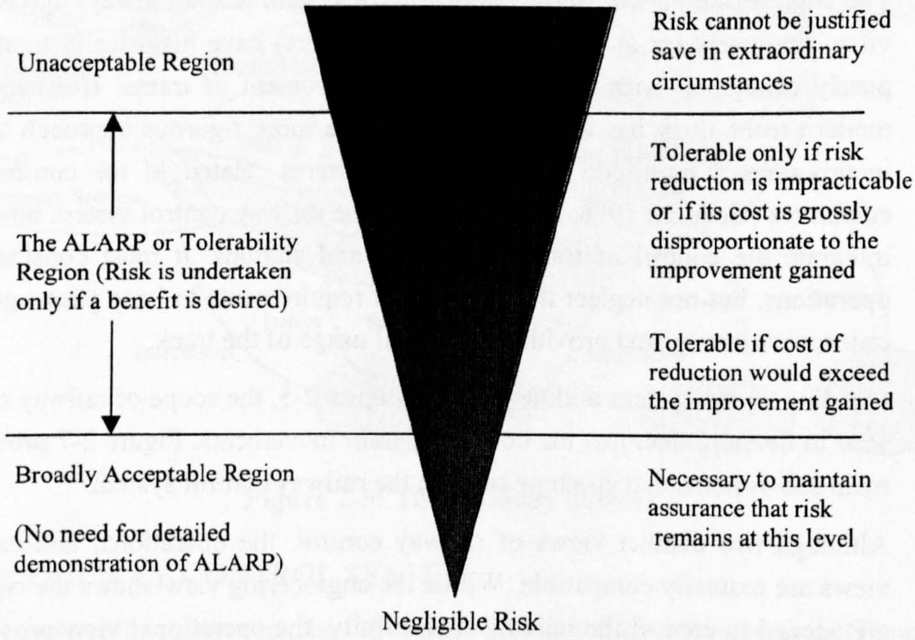


Figure 2-8: The ALARP Principle (The Engineering Council 1993, page 22)

In order to adopt this approach, the tolerability of the risk involved in any activity must be assessed. This can be done in a number of ways:

- Where risks are so high as to be unacceptable, they are often prohibited by legislation;
- Where risks are in the 'broadly acceptable' region, they can often be assessed through a qualitative risk assessment technique (e.g., 'engineering judgement');
- Where the tolerability of a risk is not so clear, quantitative risk assessment techniques can be used to clarify the levels of risk and then to compare them with the cost of possible mitigation methods to determine whether or not they are ALARP.

As the process of operating a railway exposes staff, customers and assets to a number of hazards, these must be controlled in a manner that brings the risk to individuals, corporate bodies and society in general to a level that is tolerable and maintains it at this level.

2.4.2 TOLERABLE RISK

Attempting to define the level of risk that is tolerable can be fairly contentious. The Department of Transport has suggested that people in Britain consider it worth spending £1 to 1.24 million to prevent a road death, and that this is increased to £2 million on the railways. In practice, investment rates on railways in Britain have worked out slightly higher than this, at £2.7 million per potential fatality prevented, and investment rates on road systems far lower, at £100,000 per potential fatality prevented (Davies 2000, p29; Ford 2002 p18, Harris 1997, p317).

Evidence presented to the 2001 Southall and Ladbroke Grove joint enquiry into train protection systems actually contradicted the Department of Transport position on public willingness to pay for safety improvements. This evidence showed that, even just after Ladbroke Grove, the "vast body of public opinion does not believe spending significantly more on rail safety than on road safety or health would be justified" (Muttram 2001, p5). This view was supported in the findings of studies commissioned by the Department for Transport, Local Government & the Regions, which found that after the Southall and Ladbroke Grove accidents (and the associated media coverage), a sample of individuals (40% of whome used rail at least three times a week) were only willing to pay 1.003 as much on rail safety improvements as on road improvements to prevent a fatality (Ford 2002, p18).

In line with this view, some other countries use the same value for life when assessing potential projects on road and rail (Sweden being an example) (Ring 1999, p8).

Considering the evidence of willingness to pay that is available:

- Railways spend more on safety than the public is estimated to expect
- Without counting fatalities to other road users (such as pedestrians and motorcyclists), travelling by train is six times safer per km than travelling by road (ERTMS Programme Team 2002, p10)

Thewrefore, it would be expected that the public should be well satisfied with the railways' handling of safety concerns. However, the media response to any railway accident demonstrates that this is not the case. The arguments usually put forward to explain this are that:

- The public are prepared to take higher risks when they are in control (such as when driving their car) than when they have entrusted control to a service provider (such as a railway operator), and;
- The public have a greater aversion to mass disaster.

(Burrage 1996, pA3/17).

These arguments seem fairly persuasive, until it is realised that this desire for improved safety does not transfer equally to a bus or coach operator, where the control of risk is also handed over to the service provider and ehere there is also a risk of mass disaster. Indeed, in the period from 1986 to 1995, an average of 3.3 people were killed or seriously injured for every billion passenger miles travelled on Britain's railways. The figure for bus / coach travel over the same period was 16.0, nearly five times worse (Coleman 1999).

One alternative explanation for the public preoccupation with rail safety lies in the fact that technology already exists to overcome the risks associated with rail travel in a way that it does not exist for road travel. The Bishop of Hereford, for one, has argued that this should create a higher safety requirement for railways and justifies higher expenditures (Ford 2002, p19). This view has also been supported by research into subjective evaluation of risk, which has found the perceived acceptability to be influenced by the extent to which exposure to the hazard and its potential for harm are controllable (Cox, et al. 1998, p204).

In addition, since railways represent a controlled environment, they can be fitted with secure emergency communication systems, provided with control staff to oversee traffic movements and intervene at any sign of trouble, and they can even be fitted with automatic safety systems such as ATP. Put in other words, rail safety can be improved by use of online control systems that can intervene in real time to prevent an unsafe state becoming an accident. In contrast to this, the open access environment and smaller vehicle separation on roads make the use of emergency communication to avoid accidents unworkable. There is also, as yet, no ATP equivalent solution for road traffic. The issue of road safety must therefore largely be tackled by re-modelling layouts and enforcement of traffic laws at key locations. These are off-line activities that cannot prevent an accident as it is about to happen. It is the author's view this may also be the source of a significant difference in public perception of possible measures to increase safety. Beyond expenditure, potential benefit or how preventable an incident is, public expectation is far more insistent that real time control mechanism should be implemented where they are possible than that off-line measures should be taken.

2.4.3 RAILWAY RISKS

By far the majority of fatalities occurring each year on Britain's railways are related to trespassers and suicides. Of the remaining railway fatalities, the number and relative statistical significance of any particular cause varies from year to year. Generally the order of significance is:

1. Train accidents (de-railments and collisions- including level crossings);
2. People being run-over or falling from trains;
3. Those that occur at stations (not train related).

(HMRI 1999, HMRI(2) 2000, HMRI(3) 2000 ,HMRI 2001, HMRI 2002).

Clearly, the train control system cannot be expected to make any significant contribution to the reduction of accidents that do not involve trains (such as falling down the stairs in a station). However, it is possible for some accidents in other categories to be avoided through the use of appropriate train control systems. Table 2-2 details the main accident types that occur on a railway, together with some example prevention methods. It can be seen that train control offers a number of possibilities, as do civil engineering works and rolling stock design.

Statistically, both the rate at which accidents occur on Britain's railways and the severity of those accidents are falling (Evans 1999, p886; Davies 2000, p26-27). There are a number of explanations for this trend, including: better rolling stock crashworthiness design; improved braking performance; continuing signalling system developments and the introduction of a number of driver performance improvement regimes. However, the rate of fatalities due to potentially preventable incidents of signals being passed at danger (SPADs) is not keeping pace with the general reduction in accident fatalities. In fact, the number of fatalities due to conflicting movement SPADs is rising at about 1.4% per year (Evans(1) 2000, p52)

Overall, Table 2-2 shows that the largest proportion of fatalities since 1997 were due to train accidents and collisions (47%). Considering a wider time scale, there were 76 such accidents between 1967 and 1999. Of these, 23 (30%) were due to SPADs and were ATP preventable. A further 8 (10%) were due to overspeed or buffer collisions and were also ATP preventable (Evans 1999). Table 2-3 expands on this by showing the main railway accident types, together with the number of accidents and fatalities that occurred due to each between 1967 and 1999. The figure also shows the number of accidents and fatalities of each type that were theoretically preventable by an ATP system, in total 40% of all such accidents.

The current trends in accident rates and consequences can be projected to provide estimates of railway accidents and fatalities in the future. Professor A Evans, of Imperial College London, has carried out one such study. The results of his work suggest that, if current trends continue, a total of 25 railway accidents can be expected from 2000 to 2024. Of these, 18 accidents (72%) would be ATP preventable. Professor Evans identifies the most significant cause of this increase in ATP preventable accidents to be the current increasing trend in accidents due to conflicting movement SPADs (Evans[1] 2000, p53). Other significant causes can also be found in the higher density of traffic and increased speeds now being used on some routes.

Analysis of railway safety statistics and predictions such as these show a clear potential and need to reduce railway risk in the areas addressed by ATP. In recognition of this fact, a review of UK signalling principles by the IRSE concluded that "UK signalling principles should require the provision of comprehensive train protection systems, and every opportunity to introduce train protection systems should be investigated" (IRSE[3] 2001 section R2.1).

Accident Type	Number of Fatalities	Example Prevention Methods
Train accident (Derailment, Collision with train or road vehicle)	56	1. ATP (Overspeed and SPAD prevention); 2. Enhanced emergency braking performance; 3. Better driver training; 4. Improved crashworthiness of rolling stock; 5. Provision of flyovers to replace fat junctions; 6. Driver reminder appliances; 7. Use of bridges / underpasses in place of level crossings; 8. Upgrading level crossings in use.
Fell off platform and struck by train	25	1. In cab CCTV systems on approach and whilst closing doors / departing; 2. Passenger operated emergency plungers on the platform (with risk of delays due to inappropriate use); 3. Platform attendants or monitored platform CCTV, together with a means for staff to contact drivers or initiate emergency stop; 4. Enhanced emergency braking performance; 5. Platform Edge Doors.
Struck standing near platform edge	4	1. In cab CCTV systems on station approach and departure; 2. Platform attendants or monitored platform CCTV, together with a means for staff to warn passengers, contact drivers or initiate emergency stop; 3. Removal of slam door stock from service; 4. Platform Edge Doors; 5. Controllable access to platforms on high speed lines.
Crossing lines (not trespassing)	4	1. Appropriate provision of bridges, underpasses and controlled level crossings; 2. Platform attendants or monitored platform CCTV, together with a means for staff to contact drivers or or initiate emergency stop.
Leaning out of / falling from carriage	17	1. Improved door engineering and control mechanisms, including interlocked systems to ensure that doors close before departure and then remain closed until next authorised stop; 2. Train crew Guards on Trains; 3. Reduced size of, or bars over, opening windows; 4. Emergency alarms in train carriages.
Slips, trips and falls	16	1. Non slip surfaces.
Entering or alighting from trains	6	1. Improved door engineering and control mechanisms, including interlocked systems to ensure that only correct side doors are enabled to open; 2. Guards on Trains; 6. Emergency alarms in train carriages and either on the platform or the outside of trains to warn drivers of problems / initiate emergency brake (with risk of delays due to inappropriate use); 3. Platform attendants or monitored platform CCTV, together with a means for staff to contact drivers or initiate emergency stop; 4. Reducing the gap between the platform edge and train; 5. Inter car barriers to prevent passengers falling between carriages; 6. Platform Edge Doors; 7. Standardisation of platform height and surfaces.
Struck by non railway vehicle	1	

Table 2-2: Example Accident Prevention Methods (figures for 1997 to 2002) (HMRI 1999, HMRI(2) 2000, HMRI(3) 2000, HMRI 2001, HMRI 2002) Trespassers & Suicides Excluded

Train Accident Type	Number of Incidences	Number of Fatalities	Number of ATP Preventable Incidences	Number of ATP Preventable Fatalities
Train Collision	48	174	23 (48%)	97 (56%)
Derailment, then Collision	8	15	1 (13%)	2 (13%)
Derailment	16	107	5 (31%)	20 (19%)
Buffer Collision	2	3	2 (100%)	3 (100%)
Collapsed Bridge	1	4		
Train Failure	1	1		
Total	76	304	31 (40%)	122 (40%)

Table 2-3: Fatal Collisions, Derailments and Overruns on the UK National Railway System (1967 – October 1999) (Davies 2000, pp 51-52)

2.5 RAILWAY CAPACITY

Returning to 1910, during his presidential address to the Institution of Signal Engineers, Mr Blackall stated, "in the speed and density of traffic are to be found both the necessity and the justification for a signalling system" (Blackall 1910, p12). Once again, his words are as valid today as they were 90 years ago.

Train control systems are needed in order to allow the railway to operate safely at the required line speeds and capacity. If, in order to ensure safety, the train control system restricts capacity or limits line speed unduly, it undermines its own justification. It may well be more economical to use a lower speed, lower capacity solution that does not require such high levels of control. Conversely, if a train control system can be shown to increase the safe capacity and operating speeds of the railway, then the justification for its use need not be based purely on safety – it can be shown to offer a direct benefit to the operation of the railway, with the added bonus of an increase in safety.

Whilst 'capacity' is frequently mentioned in the literature about railway systems, the term is rarely defined. This is perhaps because 'capacity' can be taken to mean a number of different things – each of which offers advantages to the understanding of a railway's performance.

2.5.1 THEORETICAL CAPACITY

Probably the most commonly used approach to measuring railway capacity within the signalling profession is that of 'headway':

> ***Headway*:** Minimum time or distance between trains that the signalling system will permit, so that a following train is not affected by the train ahead.

Measurements of capacity using this approach are most often given as 'seconds between trains' or as an inverse function of this, 'trains per hour' (tph). The term 'headway' is also commonly adopted by railway professionals of other disciplines within the railway industry, particularly in the 'trains per hour' form (see interview transcripts in Appendix A). In order to reduce the confusion that can arise from use of the term 'headway' to imply distance, time or tph, the author will adopt the following notation in the subsequent discussions:

- **Headway Distance** – the minimum distance between trains that the signalling will permit, so that the train ahead does not affect a following train;
- **Headway Time** – the minimum time between trains that the signalling will permit, so that the train ahead does not affect a following train;

- **Headway** – the maximum throughput of trains that the signalling will permit, so that the train ahead does not affect a following train, measured in tph. This term will also be used where the distinction between distance, time and tph is immaterial.

Unfortunately, the definitions given above allow two conflicting interpretations of what is being measured by headway. The common practice adopted by Main Line railways in the UK is to use the term 'headway' to describe what the author prefers to call the 'Train Following capacity' of the railway:

> ***Train Following Capacity:*** The maximum throughput at a particular point on the railway network, such as a signal position, if all trains were to follow each other at line speed and with a minimum of braking distance separation, no allowance being made for station stops.

In contrast to this, the typical interpretation of headway for UK metro railways could be described as 'Point Capacity':

> ***Point Capacity:*** The maximum throughput at a particular point on the railway network, such as a station platform, accounting for station stops and actual train speeds (McCormick et. al., 1996 pp A1/7).

In the case of both Train Following and Point capacities, the train ahead must not affect a following train for headway to be achieved. In either case, the use of headway as a measure of capacity is only meaningful in the context of a specific point on the railway. Just because one point can achieve a high throughput of trains per hour does not mean that the whole line will be able to do so.

The measure also fails to account for the speed of the service. Examples of the relationship between speed and achievable headway are given in Figure 2-9, from which it can be seen that speeds may need to be reduced to increase the achieved trains per hour. However, reducing speed on a line would require additional numbers of trains in order to provide the same service frequency and would also significantly contribute to journey time (an important part of the performance as far as passengers are concerned). Clearly, a railway operating with a capacity of 20 trains per hour averaging 60mph will be providing a much better service to its users than a railway operating with a capacity of 20 trains per hour averaging 30mph!

"It is not enough just to consider point capacities along the line. We need also to look at junctions, terminal working, services operating with different stopping patterns, trains running at different speeds and the rolling stock fleet size" (McKenna 1998, pA1/4). However, headway has its advantages in helping to identify ways in which the effects of pinch points on the network can be overcome. Further consideration is given to the subject of headway in section 2.6.1.

In order to overcome the point specific nature of 'headway', the minimum line headway can be considered:

> *Minimum Line Headway*: The minimum time interval between a pair of trains, so that a following train is not affected by the train ahead throughout its run (Pachl 2000, p5/8).

Perhaps most usefully this could be measured as seconds between trains, since it will vary for different pairings of trains. An example of how this works can be seen in Figure 2-10. If the area of line shown in the figure represented the entire line being considered, the minimum line headway between non-stopping trains 1 and 2 would be 77 seconds. Between non-stopping train 2 and stopping train 1 it would be 152 seconds, etc. Attempting to run a train over the line at less than its minimum line headway separation from the proceeding train would cause its progress to be impeded.

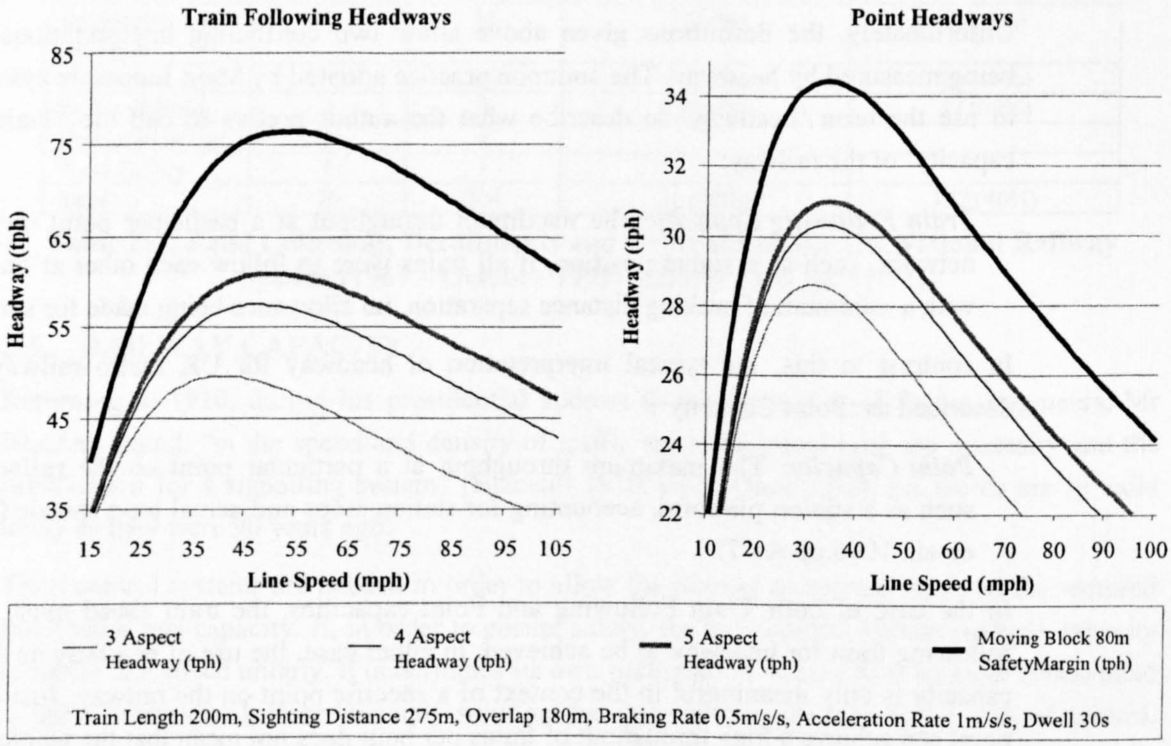
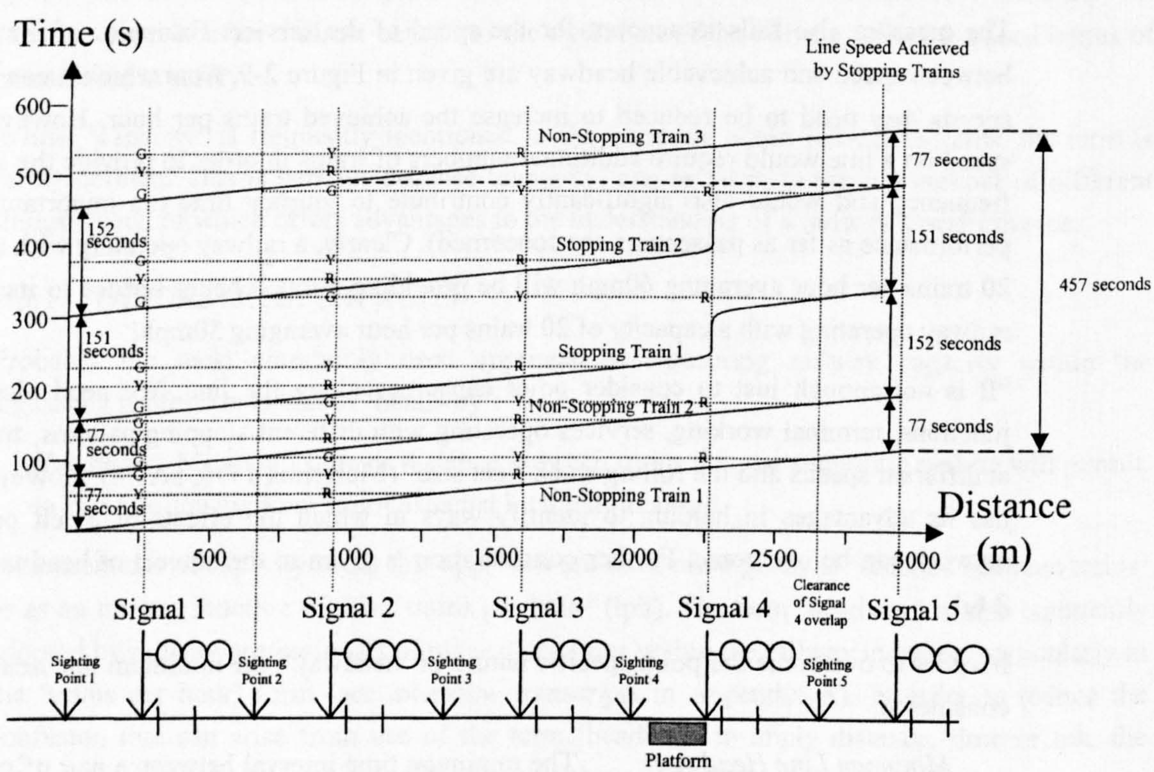


Figure 2-9: Theoretically Achievable Headway



Design Assumptions:

Sighting distance = 275m
Train length = 200m
Braking rate = -0.54 m/s/s
Station Dwell time = 30 seconds
Overlap length = 180m
Line Speed = 60mph = 25.8 m/s/s
Acceleration rate = 0.56 m/s/s

Distances and Times determined by use of Newton's Laws Of Motion.

Figure 2-10: Variation in Line Headway with Stopping Pattern

Minimum line headway can account for the combined effects of the point capacities achievable throughout the line, including the effects of all relevant junction layouts, multiple line sections, station stops and reversals. However, it only considers the headway of two defined trains and, as it considers the effect of complete journeys, it does not remove the need to consider technical headway when identifying where the system pinch points are and how to alleviate them.

An alternative definition that can be used to consider the passage of any number of trains is that of 'Theoretical Line Capacity':

> ***Theoretical Line Capacity:*** Indicates the theoretical maximum throughput of a railway line when all trains complete more than one round trip.

Appropriate units for measuring theoretical line capacity would usually be quoted as train kilometres per hour. For a defined railway line, this provides a combined measure of both the number of trains using the line and the speeds at which they travel, but does not allow these factors to be differentiated. This means that services with vastly different headways may achieve the same theoretical line capacity due to operation at different speeds.

For theoretical line capacity to be meaningful, it must represent a defined train service pattern. This can be achieved by assuming all trains to have the same nominal performance and stopping patterns (as may prove useful for a metro, but less so on a mixed traffic railway). Alternatively, some other (more realistic) theoretical service pattern can be assumed. In either case, the assumptions made and the use of the units 'train kilometres per hour' will make the measure line or network specific. The service provided by a given number of train kilometres per hour could be vastly different on lines or networks of different sizes and characteristics – or with the use of a different set of performance and service pattern assumptions.

As theoretical line capacity represents the combined effects of the engineering infrastructure that supports and enables the railway's operation, it is included in Figure 2-2 and Figure 2-3 as an output of the 'Develop Railway System' process.

2.5.2 SUSTAINABLE CAPACITY

Headway and theoretical line capacity relate to the 'theoretical maximum' that can be achieved from the railway's fixed infrastructure. However, the operation of any railway can be expected to suffer deviations from the ideal operating conditions due to factors such as variability in system and human performance, system reliability and external influences or perturbations. A margin for recovery is often added to the headway time to give a 'service interval' that allows for such deviations:

> ***Service Interval:*** The sustainable time interval between trains (Headway Time + Recovery Margin).

The service interval (usually measured as seconds between trains) is also referred to as the 'operational headway' (measured as trains per hour). Just to confuse matters further, both railway engineers and operators often shorten this to 'headway' (with the result that it is sometimes difficult to know which definition of 'headway' is being referred to).

The size of the recovery margin must be selected to account for the probability of delays occurring. The operator must find an optimal balance between the proportion of time for which the margin will be sufficient to allow sustained service and the reduction in capacity that it introduces for the unperturbed service. In line with this approach:

- the UIC have quoted a desirable norm for the actual operating capacity of a railway as 75% of the theoretical capacity during peak periods and 60% in off-peak periods (Holgate 1998, p9; Holtzer 1999, p587; Schmid[1] et al. 2002, p7);
- the SRA have concluded that the optimal usage of theoretical capacity is 75%, with higher usage requiring careful consideration (Steer 2003, p3);
- the European Commission has determined that, after accounting for the ability of the network to absorb traffic distortions, practical capacity can be 60 to 95% of theoretical capacity (Scherp 2003, p2).

Being based on technical headway, the service interval is again point specific. This limitation can be overcome by considering a derivative of the minimum line headway. This could be referred to as a sustainable line headway:

> *Sustainable Line Headway*: The sustainable time interval between a pair of trains, such that the train ahead does not affect a following train, throughout its run (e.g., minimum line headway + recovery margin).

On a metro railway (where all trains have similar stopping patterns and performance) the sustainable line headway is fairly easy to work out. However, on main line railways, with a multitude of train performance and stopping patterns to accommodate, the train separation required will vary significantly throughout the day (as already shown in Figure 2-10). In order to cope with this for planning and charging purposes, it is now common practice on UK main line railways to talk about 'train paths'.

2.5.2.1 TRAIN PATHS

A train path can be defined as:

> *Train Path:* An allocated unit of capacity on a section of line, based on the sustainable line headway between trains with a defined nominal performance and stopping pattern (Nominal Line Headway).

If the behaviour of the non-stopping train in Figure 2-10 is representative of the nominal train performance and stopping pattern (with a minimum line headway of 77 seconds and sustainable line headway of, say, 100 seconds for a 23 second recovery margin), the non-stopping trains would require one train path, whilst operation of a stopping train would require two train paths. As can be seen in Figure 2-11, some of the capacity of the line would be wasted by a rigid adherence to the allotted train path, but use of this measure provides a convenient method for determining access charges.

In practice, railways will never actually use all possible train paths. If the railway is to run efficiently whilst satisfying the users' demands, it must be expected that the service will vary throughout the day, with extra trains at peak times and limited service, or possibly even closure of the railway, due to a lack of demand during some other periods. Engineering access may also be required to carry out maintenance and renewals work. Failure to account adequately for this latter constraint would make the service unsustainable in the long term. Therefore, the service must be operated to an optimal service pattern:

> ***Service Pattern***: The sustainable timing of trains when all trains complete more than one round trip – i.e. the timetable.

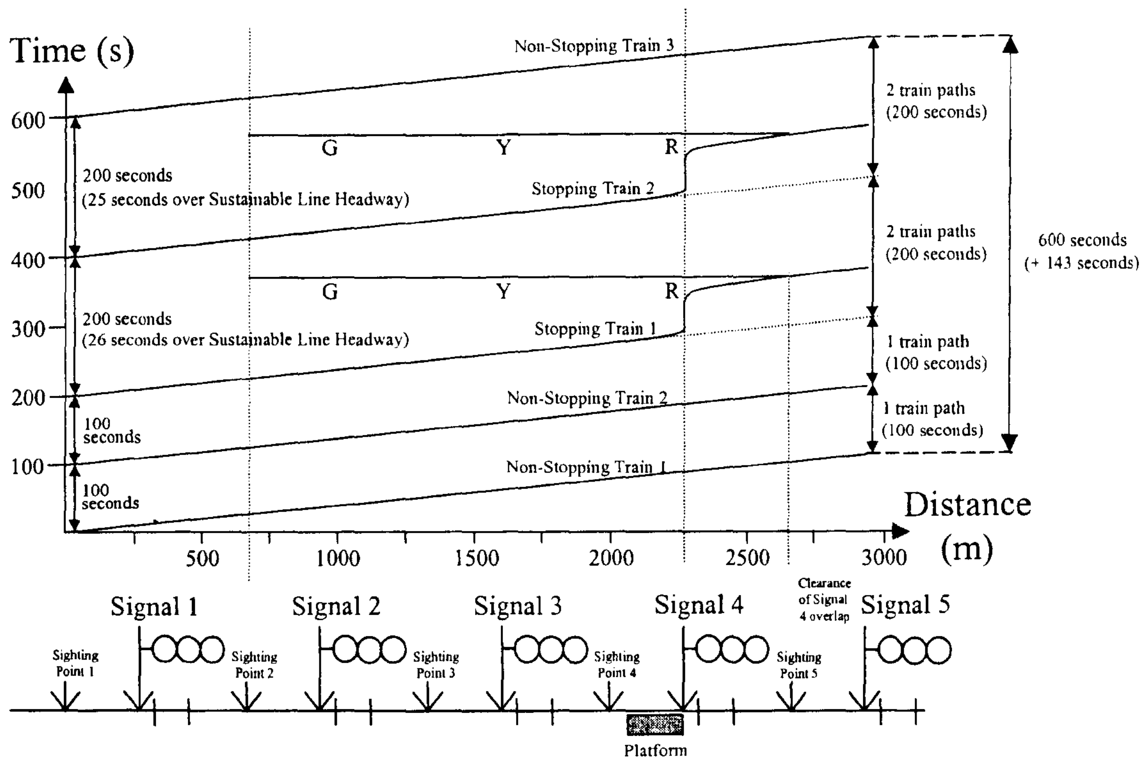


Figure 2-11: Schematic Arrangement of Train Paths

As the service pattern represents the planned capacity of the railway, it combines the effects of the engineering infrastructure and an operational margin for recovery with the potential to allow for other requirements and constraints on the railway's usage, such as:

- User demand – when, from where and to where users wish to travel, along with their desired journey duration. If the railway is to run efficiently whilst satisfying the user's demands, it must be expected that the service pattern will vary throughout the day, with extra trains at peak times and limited service, or possibly even closure of the railway, due to a lack of demand during some other periods;
- The need for engineering access to carry out maintenance and renewals. Failure to account adequately for this constraint would make the service unsustainable;
- The number of vehicles and crew needed to operate the service. This will not only determine a significant part of the railway's fixed and variable costs, but may also be constrained by stock and crew availability;
- The desire to avoid timetabled conflicting movements at junctions.

The service pattern is therefore included in Figure 2-2 and Figure 2-4 as an output of the 'Plan Railway Operation' process, an input to the 'Control Service' process and a constraint to the 'Maintain System' process.

The capacity offered by a service pattern could be defined as the 'Sustainable Line Capacity':

> ***Sustainable Line Capacity:*** Indicates the sustainable throughput of a railway line when all trains complete more than one round trip, in accordance with the time tabled service pattern (based on McCormick et. al., 1996 pp A1/7).

Again, the units of this measure could be train kilometres per hour. However, it should be noted that neither the measure of trains per hour, nor that of train kilometres per hour considers the loading capacity of the trains themselves. Clearly, the capacity of a railway operating 20 train kilometres per hour with 4 car trains will only achieve the same passenger throughput as a railway operating 10 train kilometres per hour with 8 car trains. The capacity of the actual trains themselves is therefore of great significance to the overall capacity of the railway. Accounting for this, the units of our theoretical line capacity and sustainable line capacity could be converted from train kilometres per hour to Passenger/tonne kilometres per hour.

All that is required to convert train flows into passenger / goods flows is an understanding of vehicle loading. This is a relatively simple matter on railways where all trains are the same, or have broadly similar characteristics (such as a typical metro railway), as long as the measure of capacity refers to theoretical maximum capacity (i.e. assumed full loading). However, if the measure is taken to be actual loading (which must then be measured or estimated in some way), the conversion becomes more difficult. This is particularly true if the railway operates mixed stock / traffic types with long and short trains able to carry differing loads of passengers or goods. As a consequence, and in order to simplify the issue, most papers referring to line capacity do so in the context of train numbers only.

As the sustainable line capacity is based upon the theoretical line capacity, and its measure is based upon train kilometres per hour, it does not overcome the limitations already outlined. It simply provides a more realistic measure of what can actually be reliably achieved by a railway.

2.5.3 ACHIEVABLE CAPACITY

It is important to note that all of the capacity definitions above are actually theoretical, as there will always be the potential of disturbances occurring in the system for which the capacity calculations did not allow. In reality, railway capacity is a very complex issue, with numerous engineering systems required to work together with each other, within the context of the railway environment, in order to deliver the resulting capacity. Building on the IDEF0 system model and capacity definitions already discussed in section 2.2.1, the main factors that contribute to the real capacity achieved by a railway system are shown in Figure 2-12.

In practice, in order to achieve a given 'Achievable Line Capacity', the railway system must be designed for a higher 'Sustainable Line Capacity' and a higher still 'Theoretical Line Capacity'. The 'Point Capacity' or 'Train Following Capacity' required to achieve this level of service will then vary through the line depending on service patterns.

It should be noted that the system's capacity can be drastically altered by the physical constraints of the infrastructure – both the track and the rolling stock. For example, the track clearance gauge will determine the maximum size of the train carriages, which will in turn limit the capacity of that carriage. Similarly, the train performance (acceleration and braking rates achievable by the rolling stock, together with their maximum speed) will have a major effect on the achievable headway.

In practice, capacity within a given system is achieved through a series of trade-offs in system characteristics. For example, the internal design of rolling stock will have an effect on capacity through the number of seats that are provided. Fewer seats mean that more people can fit into the carriage as standing passengers. This effectively increases the carriage capacity and makes more space available for doors – both of which improve the quality of service provided in demand peaks (Glover 2000, p54). However, if too few seats are made available, large numbers of off-peak passengers may be forced to stand as well, which reduces the quality of service for them (McCormick et al. 1996, pA1/9). Similarly, as far as the control system is concerned, it may be possible to run a higher capacity service at slower speeds if the advantages of reduced platform crowding are seen to outweigh the resulting increase in passenger journey time.

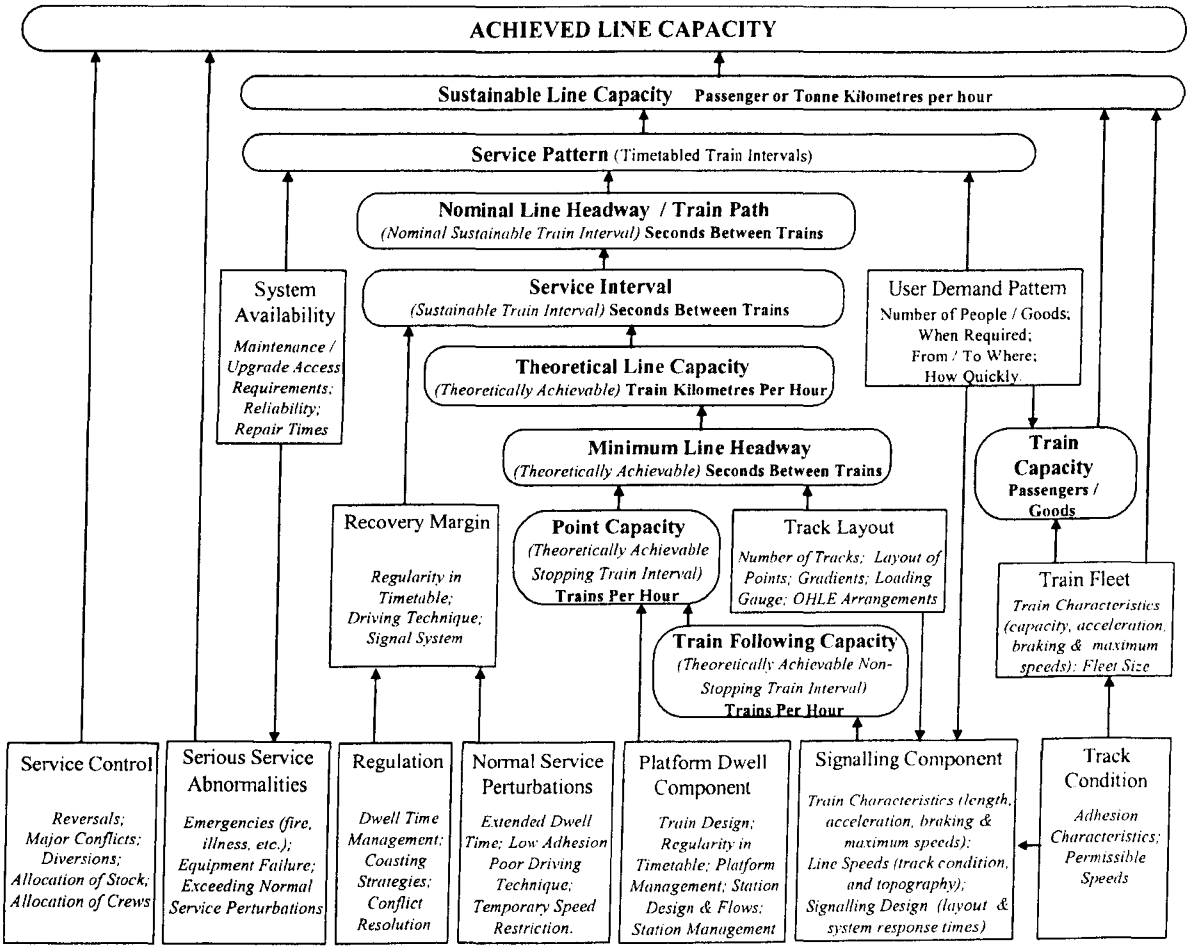


Figure 2-12: Achieved Line Capacity

A further point to note from Figure 2-12, is that the control system must be able to cope with significant perturbations from both internal sources and external sources outside of its control. This is usually done in one of two ways:

a) recovery margins built into the service interval to allow the system to cope with normal service perturbations;

b) facilities for service control to allow the system to cope with the serious service abnormalities.

It can be seen from this that the required capacity cannot be achieved by any one part of the system in isolation.

2.5.4 OPTIMISED CAPACITY

The best achievable capacity for any line or network has been defined as the theoretical line capacity. This could, therefore, be considered to be the ultimate engineering target for optimisation of capacity. However, simply improving the theoretical line capacity does not guarantee that the capacity has been optimised. The usage of available capacity must also be considered. To this end, capacity efficiency could be defined in operational terms as:

$$\text{Capacity Efficiency} = \frac{\text{Achieved Line Capacity}}{\text{Theoretical Line Capacity}}$$

Improvements could then be made to this ratio by developing the availability of the railway's systems and fine-tuning its management and control processes in order to reduce the required recovery margin, reduce service perturbations and allow better response to any perturbations that do occur. Operational measures within the service pattern, such as flighting of trains with similar stopping patterns and performance, would also improve this measure.

Unfortunately, these approaches consider only the output of the system (the capacity that can be delivered) but not the input (the transport demand and / or railway capacity demand of Figure 2-1 and Figure 2-2). As a result of this, they fail to measure the success of the system in meeting the demand that it exists to fulfil.

2.5.4.1 CUSTOMER FOCUSED CAPACITY

Optimisation from the customer perspective may well be different from optimisation from an engineering one. User requirements for the service (the required frequency of trains, from where, to where, and how long the journey should take) must be taken into account when considering the capacity achievable by a railway system. The users' requirements can then drive both the development of the system to achieve higher capacities where needed and also the optimisation of the way the railway is operated once the system has been designed. Whilst the capacity definitions considered so far have allowed for this (through the inclusion of user requirements in the development of the service pattern), they have not provided a means of measuring the success with which user requirements have been met.

In order to provide a better measure of how well the system meets the users' requirements, a further definition of capacity can be given as the 'optimum line capacity':

> ***Optimum Line Capacity:*** Indicates the sustainable throughput when passenger / goods travel times and comforts are optimised (McCormick et. al., 1996 pp A1/7-10).

'Optimised' here does not necessarily mean minimised / maximised, since the cost implications of that could be far from optimal. Instead, it implies a cost / benefit consideration of the service provided to the passenger. Suitable units for this would be social benefit value per passenger kilometre, where social benefit represents the value of the passengers time / comfort (including both journey and waiting times). This value could be optimised in accordance with achieving:

$$\text{Benefit to Cost Ratio} = \frac{\text{Social Benefit Value} + \text{Revenues}}{\text{Cost of providing Service}} = \text{Acceptable Ratio}$$

The social benefit offered by the system would be difficult to measure and, indeed, the author has not come across any published attempt to do so in a way relating specifically to customer demand. One approach that has been used on LUL is an estimation of the value of an average passenger's time, together with effects of proposed improvements on journey time and quality (LUL 1994, pp6, 9). This can even be taken so far as to consider the effects of delays in different parts of the system, such as at platforms, on trains or whilst travelling to the platform (Millard 1999, pp142-3). Such analysis provides the opportunity to tune the service to customer requirements, rather than purely technical ones.

The use of benefit to cost ratios is considered further in section 2.7.

2.6 RAILWAY CONTROL SYSTEM CAPACITY

If the capacity hierarchy shown in Figure 2-12 is considered, it can be seen that there are five components through which a railway's control system influences achievable capacity: The signalling; Control of platform dwell; Service control; Regulation and System availability.

Of these, the signalling, service control and regulation components are controlled through the train control system, and the platform dwell component can be controlled by a combination of train control and station control systems (including manual systems). System availability is not so much controlled, as a limiting factor, applying to all aspects of the railway system, including the control system.

2.6.1 HEADWAY

As shown in Figure 2-12, for a given train control system, the signalling and platform dwell components combine to define the theoretical point capacity, or headway. The relationship between headway, line speed and the train control system's signalling strategy has already been considered in section 2.5. Therefore, that discussion will not be repeated here. However, signalling strategies are not the only component of a train control system that influence achievable capacity.

Table 2-4 shows typically achievable point capacity headways for different types of train control systems on a metro railway (the details of which are described in chapters 4 and Appendix C).

Train Control System Type	Typical Achievable Headway (tph)
Conventional trainstop system, manual driving	28
Conventional trainstop system, manual driving and speed control	31
Coded ATP system with manual driving	31
Coded ATP system, ATO driving	34
Moving block transmission based system, ATO driving	36/37

Table 2-4: Achievable Headways (White 1998, p3).

It can be noted from Table 2-4 that the introduction of speed control or full ATP increases the line capacity by around 3 trains per hour, largely due to the reduction in overlap lengths that is possible once control of the train's approach speed can be enforced. The introduction of ATO to the system also provides an improvement in achievable headways of around 3 trains per hour. This is largely due to the higher consistency in driving performance that automatic operation allows, which in turn reduces the number and size of minor perturbations that occur to the service and enables the use of smaller recovery margins.

Similar headway improvements have also been found to occur on main line railways as advanced control and automation systems have been introduced. One such example is the Milano-Napoli and Torino-Venezia lines in Italy, where a reduction was achieved in the required recovery margin from 25% to only 5% of the running time, through the introduction of centralised control systems, ATO and modified track layouts (Fairbrother et al. 1993, pp116-7). Studies on behalf of NS Reizigers have also suggested that a 10-25% increase in line capacity should be possible in the Netherlands if a modern ATP/ATC system were to be used on their existing infrastructure (Holtzer 1999, p587).

This shows both that the capacity of a railway depends significantly on the type of train control systems in place, and that there is significant scope for capacity improvement on UK railways.

2.6.2 *OPERATIONS CONTROL*

The provision of a recovery margin within the service interval provides an ability to maintain capacity despite minor service perturbations. However, this ability must be managed in order to ensure that the promised 'recovery' does occur without serious disruption to the scheduled service. The systems and processes to provide this management of minor perturbations and control to the timetable are known as 'regulation'.

Where more serious service abnormalities occur, it is also necessary to manage the service in real time to ensure that train destinations are appropriately balanced, that bunching / conflicts are minimised, and that staff and stock resources are available when and where required. This function is referred to as Service Control.

The combined functions of regulation and service control are referred to as 'operations control'.

The structure of an ideal operational control system is outlined in Figure 2-13. On main line railways in Britain, the functions required within this are currently divided into a number of staff roles:

- **Signallers** deal with normal running and the duties of signalling supervisors where they are not provided;
- **Signalling supervisors** deal with local perturbation management, such as re-routing trains from fast to slow lines and changing the planned train order at junctions;
- **Station Supervisors / Managers** deal with station control activities, such as re-platforming trains;
- **Infrastructure maintenance controllers** deal with infrastructure (signalling, track, etc.) maintenance and faulting activities;
- **Stock controllers** deal with rollingstock faults, re-allocation of stock and ensuring availability of stock for maintenance;
- **Train crew resource supervisors** deal with incidents requiring changes to crew allocation;
- **Train running controllers** deal with more serious perturbations by re-timing, cancelling, diverting or terminating short of destination any affected train on a given line or geographic area.

These staff roles may be divided geographically (perhaps line by line) and split between different organisations, Returning to the example of British main lines, the signallers and train running controllers would be a part of Network Rail, the stock controllers, train crew supervisors and station managers would be a part of the train operating companies and infrastructure maintenance controllers a part of Network Rail or an infrastructure maintenance company.

A survey of European railways in 1993 revealed that there are very few automatic solutions to dealing with serious perturbations, such as signalling failure (IRSE 1993 p15). However, "the effectiveness of control depends primarily upon the quality and timeliness of the decisions made" (Day 1990, p5). With modern computer techniques there is an opportunity to consider the provision of enhanced decision support tools (such as simulators running faster than real time) or even the automation of some, if not all, of the service control functions (Schmid 2002, p11). Whilst this is beginning to occur in European railways, it is as yet a vision for the future in the UK. To date, the closest UK railways have come to this ideal is arguably the use of train graphs to assist signallers at the Cheriton Eurotunnel and Ashford CTRL control centres, or perhaps the 'Control Centre of the Future' installed in several signal boxes on Network Rail infrastructure that provides controllers with up to date train running information for the whole network. Other initiatives, such as the LUL Central Line ATR and the now cancelled network management centre for WCML, have been launched with the aim of providing improved network control (including predictive simulation for 'what if' scenarios), but these have yet to come to fruition. Clearly this is an area where there is room for improvement in future railway control systems.

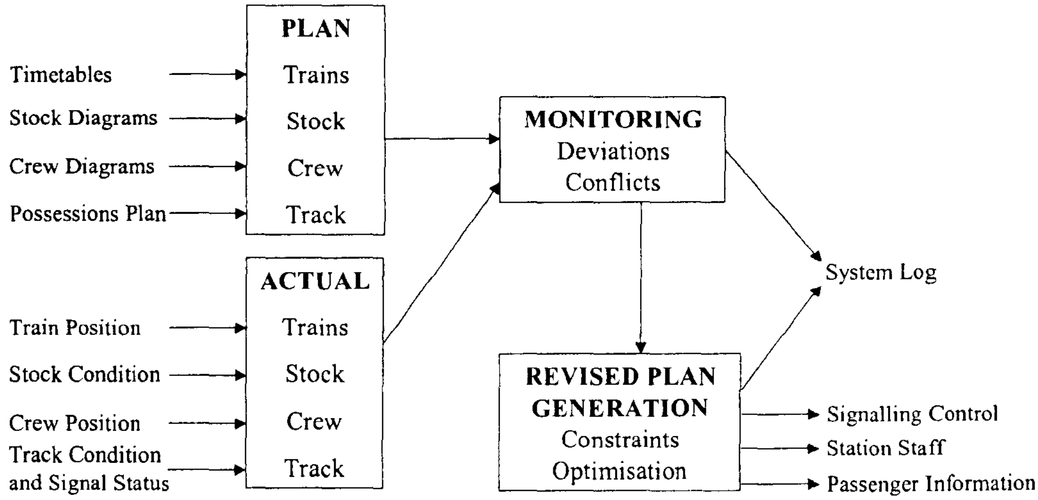


Figure 2-13: The Ideal Operations Control System (based on Hurley 1999, p132).

According to the Strategic Rail Authority, four of the top five causes of poor performance on Britain's main line railways are related to operations control. These are: operational management (the real-time operational decisions needed to recover from incidents); realism of the timetable (whether the timetable reflects actual operating constraints – such as dwell times); planning for recovery (the availability of recovery time in the timetable, spare rolling stock & crew) and congestion (recovery following an incident becomes more difficult as the proportion of theoretical capacity that is actually scheduled for use in a given area is increased). The fifth significant element is defined as asset condition (the reliability of rolling stock and network infrastructure), which will be discussed in section 2.6.4 (Steer 2003, p2).

In all, about 35% of delays experienced on Network Rail infrastructure could be classified as being due in part to technical timing rules, which could be addressed through better management of the allowances used in timetable planning (Cooper 2003, p4).

2.6.3 PLATFORM MANAGEMENT

Reducing the time that a train is stationary in a platform (the dwell time) is a major concern for both main line and metro railways that operate with a high traffic density. Dwell times result from a number of delays associated with train and platform design, service regularity, operating practice and passenger behaviour, the effects of which can be limited by implementing systems and techniques for platform management:

- **Door opening:** Enabling door opening, passenger operation of opening devices and actual door movement generally takes two to three seconds (Howarth 1999, p157). It is possible to gain reductions in this delay by releasing the doors at slow speed (as use to be possible with hand operated doors on LUL and is still the practice on Paris Metro). Research on LUL's Victoria Line has also shown that poor stopping accuracy with ATO has significant effects on door opening delays, due to the need for driver intervention if sufficient accuracy has not been achieved (Horsey 2000, p52);

- **Alighting/boarding**: Passenger alighting and boarding times are largely determined by the number and distribution of passengers within the train and along the platform; the size of any gap between the platform and train; the number and size of doors and vestibules on the train; the size and egress capability of the platforms (Adeney 2001, p2; Barter 2001, p9; Howarth 1999, pp157-8; McKenna 1998, pA1/4). Public address announcements, audible warnings as doors are about to close and playing military style music have all proved effective for reducing these delays (Harris et al. 1992, pp53-4; Horsey 2000, p22; McKenna 1998, pA1/4). Elsewhere, providing platforms on both sides of a train to segregate boarding and alighting, measures to distribute passengers evenly along the length of the train, platform queuing, additional platform staff attendance and a heavier door closing force have also helped (Glover 2000, p78; Horsey 2000, p23). The number of people congregating on the platform also increases with the time between trains, causing extension of alighting and boarding times (Horsey 2000, p72). The capacity and reliability of the train control system are therefore also of significance.
- **Door closing:** Studies have shown that operating trains with guards reduces door closing delays, whilst use of 'Close Doors' and 'Right Away' signs activated by station staff increases them (Howarth 1999, p158). Reductions can also be gained by the introduction of countdown clocks to advise drivers of time to departure and the development of an 'every second counts' culture amongst staff (Horsey 2000, p23). Of course, door-closing delays may not be solely due to drivers and station staff. Failure to clear the starter signal (whether through staff oversight or system delays) and passenger action (holding the doors open for a friend) can also contribute significantly to delays. The former can be improved by addressing staff training and culture, or by amending system algorithms. Some remedies for the latter have already been considered under alighting / boarding.
- **Traction application**: DB in Germany have reduced the delay between door closure and start away by allowing pre-selection of the traction system, so that trains automatically start away when the doors are closed (Howarth 1999, p159).
- **Turnaround:** Where a train must depart from a station towards the direction from which it arrived (such as at a terminal station), the leading cab must be 'shut down' and the trailing cab activated prior to departure. If a single driver has to perform these tasks, he/she must also walk between the cabs. This typically results in a dwell of between 6 and 10 minutes. If a second driver is available to prepare the trailing cab and then take the train out of the station, the turnaround dwell can be reduced to between 2 and 3 minutes (Barter 2001, p9).

A quick glance through the above breakdown of constituents to dwell times reveals the influence of passenger management systems, train systems, signalling and control systems, drivers, signallers and station staff. All without even considering the influence of run in/out times on the wider issue of platform occupancy.

Interestingly, most of the systems and techniques for dwell time reduction outlined in this section can be implemented without the need for major infrastructure changes. Platform management therefore offers potential for optimising the utilisation of capacity in both new and existing railway control systems.

2.6.4 AVAILABILITY

Whilst improving the service control functionality would enable the railway to cope better with service abnormalities, it would be far better to avoid the occurrence of those abnormalities and their consequent disruption to services all together.

In May 2000, Railtrack's Performance Director wrote that a major source of overall delay to the Railtrack network was from "problems with reliability of infrastructure" (Curley 2000, p49). Indeed, in 1999 responsibility for around 40% of all railway delays in the UK were allocated to the maintenance contractors responsible for maintenance, repair, minor renewals and 'rapid response' faulting for the fixed infrastructure contracted to them (Winder 1999 p 2). It is therefore clear that the availability of the railway infrastructure represents a major factor in the overall capacity that a railway can achieve.

The availability of the railway control system could be improved in a number of ways, including:

- **Improved maintenance** - it has been estimated that the combination of poor specification and inadequate execution of maintenance accounts for 64% of all infrastructure related delay minutes (Curley 2000, p49)
- **Improved equipment access** – most railway infrastructure is currently located near to the track, making maintenance access difficult. Future design of control systems must keep as much of the equipment as possible out of the track environment would greatly assist maintainers in gaining adequate maintenance time (Winder 1999 p3)
- **Improved equipment reliability** – this could include the design of components to reduce fault rates or the use of fault tolerant design such as active redundancy (with multiple sub-systems operating concurrently), standby redundancy (with an operating sub-system and one or more back up systems available to be switched in if a failure occurs) or graceful degradation (where the system is designed to continue operation at a lower level of functionality in the case of a sub-system failure) (IRSE 1993, p5, 6, 9)
- **Improvements to equipment repair times** – use of condition monitoring equipment and diagnostic tools to assist maintenance staff, together with modular designs for ease of repair / replacement (Winder 1999, p4)
- **Rule changes** – railway operating rules tend to impose a heavy penalty of delay that is often disproportionate to the original failure. For example, delays due to a failed track circuit could be significantly reduced if, rather than requiring all trains to be stopped and cautioned, the rules permitted the use of suitable signage to inform approaching drivers that a signal is 'out of use' and advise an appropriate speed restriction (Nichols et al. 1996, p37; Winder 1999 p4).

It is worth noting that the advantages of improved availability go beyond capacity benefits. For example, existing railway control systems tend to be designed to be 'fail-safe' by stopping all train movement, leaving the human operators to continue running traffic in accordance with rules and human centred control processes. This transfers safety responsibility from the technical systems to the human operators and, in so doing, introduces additional risk to the operation of the railway. In fact, it can generally be said that in the railway domain "the lower the system availability is, the lower the system safety is" (Boycott et al. 1999, p1). The provision of alternative systems designed with inherent fault tolerance and ability to operate in degraded modes would therefore not only increase the availability of the system and its capacity during times of technical system failure, but would also provide a means for the system to maintain safety besides reliance on the human element (Francis 1994, p1).

2.7 FINANCIAL APPRAISAL

In section 2.4 the financial assessment of potential safety measures against a 'cost of a life' valuation was considered in order to determine whether a risk was ALARP or required implementation of an appropriate mitigation measure. This technique has now become one of the main financial appraisal methods for determining where investment should be targeted. Unfortunately, this technique alone has serious weaknesses, in that safety is not the only benefit to be offered in return for investment.

In his paper 'signals for the future', Mellitt concluded "with the new industry structure, whole life costs will begin to dominate the approach to system specification and procurement" (Mellitt 1996, p8). This view was expanded in a 1998 paper to the IRSE, pointing out that whole life costs include maintenance and adaptability to changing business requirements. As a result, redundancy, the provision of degraded modes, training requirements and design support must all be considered when assessing potential systems (White, et.al. 1998, p8).

The problem is to develop a financial framework which allows safety, commercial and other objectives to be addressed simultaneously (Brearley 1993, p98). There are numerous methods in use, or proposed, for implementation of the ideal financial appraisal. Most of these could be classified under the general title of 'cost benefit analysis' (CBA).

CBA attempts to provide a means of analysing projects that might not immediately appear to be financially viable. It does this by calculating a ratio of benefits to costs:

$$\text{Benefit to Cost Ratio} = \frac{\text{Revenue + Other Benefits}}{\text{Capital + Operating Costs}}$$

The 'operating costs' element of this equation refers to whole-life costs, whilst the scope of 'other benefits' may vary according to the interests of the appraiser (Harris 1997, p27). Hence, a regulatory authority may consider socio-economic benefits, whereas a railway operator will be more interested in the direct benefits that they stand to receive (such as reductions in accidents). In both cases, projects offering reduced life-costs and additional benefits are likely to appear more attractive when assessed by CBA than they would by the cost of life approach. Similarly, a project is more likely to gain approval through a CBA type assessment if consideration has been given to the whole life-costs of the systems that it will implement, together with how any potential additional benefits can be best optimised and utilised.

As an illustration of this, it has been noted that the value of £2 million per life used to assess the tolerability of risks on a railway in the UK does not properly assess all of the costs associated with rail accidents. These would in practice include damage to the train, direct revenue lost due to passengers being diverted or not travelling, the cost of emergency infrastructure repairs and indirect revenue loss due to loss of confidence in the railway. Such costs can amount to £25 million for a major accident (Harris 1997, p27). Considering that there were 31 major ATP preventable accidents between 1967 and 1999 resulting in 122 fatalities (Table 2-3) this equates to an accident cost of up to £775 million, or £6.3 million per life. Predictions for 2000 to 2024 suggest an even higher value, where 18 ATP preventable accidents resulting in 3.5 fatalities per accident (a total of 63 - Evans(1) 2000, p53) could be expected to cost the rail industry £450 million (equivalent to £7.1 million per life). If it is considered that these figures do not account for positive 'additional benefits', such as increased capacity and the potential for reduced trackside infrastructure, there is clearly scope for justifying even higher expenditure than this.

It is therefore clear that the combined train control systems should be designed to optimise the ratio of capital and whole life cost against revenue and other benefits, including safety, capacity and reliability improvements.

2.8 SUMMARY

In this chapter the author has considered the environment within which train control systems operate, their purpose and their main requirements and constraints. In the process it has become clear that the train control system is central to achieving the safety and operational requirements of the wider railway system.

Several key areas of interest in optimising the train control system's performance have also been identified. In particular:

- The system must address both of the main requirements of maintaining safety and providing operational capacity;
- Predictions of the risks associated with railway operation show that a significant improvement could be made by the introduction of ATP, particularly in mitigating the increasing risks of conflicting movement SPADs;
- There would appear to be scope for capacity improvements through the introduction of centralised control and automatic control systems such as comprehensive ATP and ATO;
- If the required capacity is to be achieved consistently, the train control system must be designed with inherently high availability for operation of the railway. This can be improved in a number of ways, including component design, overall system architectures, and design for maintainability;
- Railway capacity is the product of a complex combination of systems and processes. If the capacity is to be optimised the railway must be treated as a system, with all of the factors contributing to achieved capacity accounted for in the design of its control systems.

Public opinion demands that railways implement control systems that may not be justifiable under a 'cost of a life' approach. However, when the wider cost benefit analysis is considered, there is considerable scope for providing improvements in safety and levels of service that would meet the aspirations of both the railway operators and their customers.

3 DEVELOPMENT OF 'SIGNALLING' BASED TRAIN CONTROL

3.1 INTRODUCTION

Whilst the purposes of train control systems have not changed over the years, the ways in which these purposes have been achieved have undergone numerous changes, mainly in response to accidents and advancing technology. In this way, the discipline of railway signalling has evolved over the last hundred years or so. "Arguably, it has never been founded on pure theory or philosophy and its development has been driven by the need to extract greater capacity from the railway infrastructure" (IRSE[(1)] 2000, p2).

In this chapter, the author considers how the primary requirements and hazards of railway operation have been dealt with historically by implementing a combination of rules and technological solutions within a signalling system. It is intended that this will help the reader to develop an understanding of the railway control systems currently in use and that it will also demonstrate that there are a number of ways of solving the problem of train control.

Railway signalling represents a part of the overall railway control system. It embodies the same purpose as the wider train control system, namely to ensure that trains run safely and efficiently (see section 2.1). However, its scope has traditionally been focused upon the technical means of maintaining a safe separation between trains through the use of wayside equipment. As a result, the terms 'safety' and 'efficiency' have been given a narrower meaning. Hence, according to the IRSE the fundamental safety requirements of a signalling system are "keeping trains adequately separated from each other, and stopping (or slowing) trains where necessary to avoid potentially unsafe situations", whilst they define efficient use of the railway as "mainly about using the minimum of infrastructure to provide as many train paths as possible" (IRSE[(1)] 2001, paragraph 30).

In order to achieve the purposes of this scope, the main functions of a signalling system can be considered as:

- **Lock:** to set up a safe route for the passage of each train over the track that it is to traverse, preserving the route in front of the train whilst it is making its movement;
- **Block:** to maintain a safe separation between trains;
- **Interlock:** to prevent conflicting moves at junctions, when crossing a route that could be taken by another train and at level crossings;
- **Unlock:** to release the route (for use by other trains) after the passage of the train.

(Goddard 2002, pp1-2; IRSE[(1)] 2001, paragraph 31; Taskin et al. 1995, p1).

These functions represent those activities required for route availability and integrity. In order to utilise them, the signalling system must also have a fifth function of issuing movement authorities, that is means of authorising a train to enter a route route that has been set up for it and of indicating the maximum safe speed relative to track, geometry and distances to signals or obstructions, to the driver.

The communication of movement authorities to train drivers has traditionally been achieved through signal aspects at fixed locations along the track, supplemented by driver route knowledge and speed limit indication signs. However, in recent years the technology has become available to achieve these aims by other means (such as inductive and radio communication to in-cab displays) and also to supplement the movement authority by supervision and enforcement. The distinction between signalling and train control has therefore become blurred. Hence, the IRSE also define a sixth function of signalling as:

- To "supervise and / or enforce the train to stay within its movement authority" (IRSE[(1)] 2001, paragraph 31).

A seventh function can also be identified in the majority of current signalling systems. That is, the ability to detect and protect against some failures of, or damage to, structures, track or railway formation (Goddard 2002, p2). This function has in part come about as a by-product of the technology used in signalling systems. This is particularly true of track circuits, which provide a convenient operational means of replacing signals to danger from the location of an incident, This can be by use of track circuit operating clips, or in some instances automatically - e.g. following some rail breaks or obstruction of the line by conductive material. In other respects, such as the provision of control functions for replacement of signals to danger when the need is known, this function is more fundamental to the purpose of the system. However, even the by-product functionality has become widely relied upon to provide protection under some circumstances and, therefore, it must now be considered as an element of signalling systems, particularly when updating them in ways that removes the traditional protection mechanisms.

When asked during interviews by the author to define the difference between signalling and train control, professionals experienced within the fields of railway signalling and operations did not all agree (see Appendix A). Between the views expressed during interviews and those found within a literature review on the subject, the author identified one common definition of signalling but five general definitions of train control, as shown in Table 3-1.

<table>
<tr><th></th><th>Definition of signalling</th><th>Definition of train control</th><th>Source References</th></tr>
<tr><td>1</td><td rowspan="6">Signalling is the use of fixed trackside or in-cab signals to give a driver (whether human or an ATO system) an indication of the point that it is safe to proceed to.
Signalling equipment includes train detection, interlocking, point machines, level crossings and colour light signals / in-cab signal displays. Signalling also includes the organisation and regulation required to ensure the safe and efficient use of signalling equipment.
Signalling does not control the train at all. The driver does that based on interpretation of the signals. It therefore provides a safe environment (preventing trains hitting each other or being derailed) only if the driver correctly follows the signals.
Signalling can exist without any control of the train itself.</td><td>Train control deals with all aspects of the control of train movements for the safe and effective operation of the railway. This includes regulation, control of the train itself, traction and rolling stock maintenance scheduling and diagramming, train crew management and centralised passenger information dissemination. Train control includes signalling functionality and my also include automatic functions for train protection and operation (ATP / ATO).</td><td>State of The Art 1998, p341-2; See interviews (D Hayward; B Hills, J Lewis, D McKeown, G Poitrasson-Riviere; A Rowbotham)</td></tr>
<tr><td>2</td><td>Train Control is a part of the signalling system (role not defined).</td><td>Taskin et al 1995, 9A3/p1,9</td></tr>
<tr><td>3</td><td>Train Control deals with strategic and tactical management of traffic and the railways resources. This includes the control of access to sections of the infrastructure.</td><td>Schmid 1999, p5; See interviews (Sir D Davies)</td></tr>
<tr><td>4</td><td>Train control is the local control of a train, including advising the driver (whether human or ATO) about features that support efficient running.</td><td>See interviews (J Carpenter)</td></tr>
<tr><td>5</td><td>Train control is to do with supervision of the actions of a driver. It would be superimposed on the signalling system.</td><td>See interviews (P Hosey)</td></tr>
<tr><td>6</td><td>No definition given</td><td>British Standards Institution 1998, p1,2; Khessib 1989, p2-- quoting a US congress report (1976); Nishinaga 1994, p2; Pope 1975, p5</td></tr>
</table>

Table 3-1: Definitions of Signalling and Train Control

Having considered these differing definitions, the author has decided to adopt the majority view, that train control includes all aspects of the safe and effective control of train movements. A part of this, namely the control of trackside infrastructure to set up and maintain a safe route (including providing the driver with an indication of the point to which it is safe to proceed) is often referred to as signalling. Under this definition, the introduction of train supervision and / or protection moves beyond the pure field of railway signalling and into the wider area of train control. In accordance with this approach, Figure 3-1 provides a representation of the traditional signalling based train control system and the part that signalling plays within it.

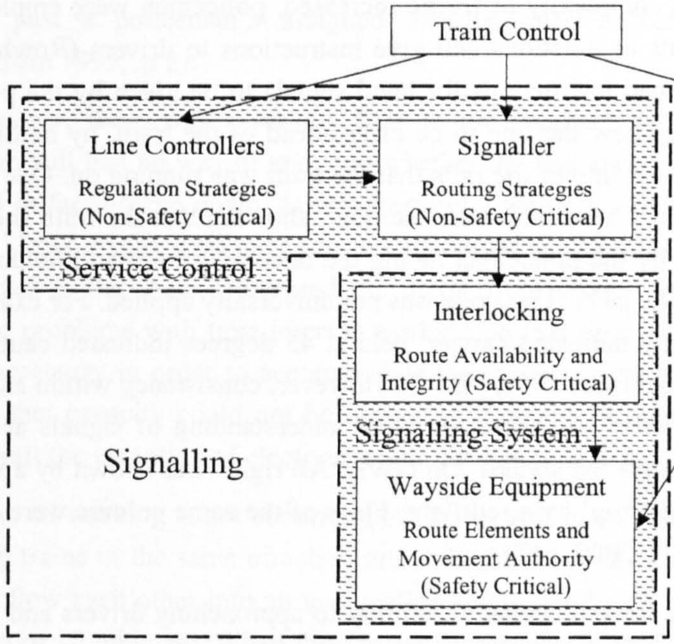


Figure 3-1: Signalling Based Train Control System Hierarchy

As signalling systems have developed, so have the procedures within which they operate. At various times in their development the split of responsibilities between the engineering functions of railways signalling and the systems operators have varied. In this chapter, the author considers the development of railway signalling, together with the involvement of humans in operating the various systems that have resulted. In so doing, he sets the scene for a subsequent investigation of more recent approaches to the subject of train control in Chapter 4.

3.2 ON-SIGHT WORKING

Early railways had their origins in collieries, where horse drawn wagons operated for many years on a system of 'drive by sight'. Whilst speeds remained low and the numbers of trains few, this type of system worked very well (Hall 1996).

A 'drive by sight' control process relies upon a combination of the human driver's geographical knowledge and visual observation. This is in effect a closed-loop control system, as shown in Figure 3-2. Whilst operating under this type of system, observation of the desired movement target and the feedback loop determining the error that exists between that target and the trains actual movement are provided by the driver, not the system.

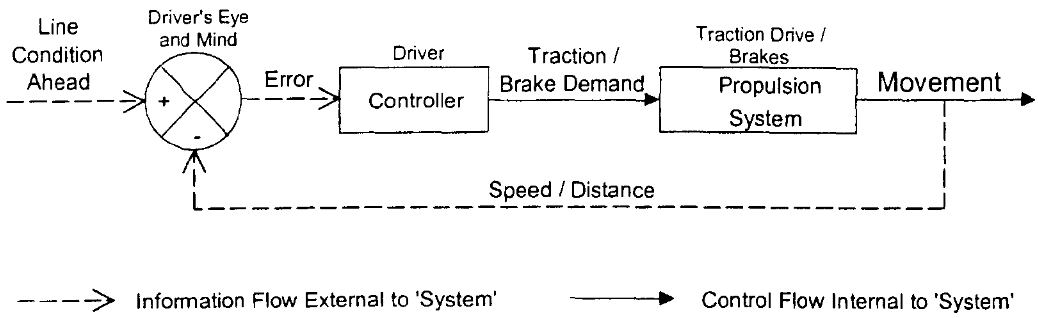


Figure 3-2: Basic Human-Machine Train Control System (see Appendix B)

As the density and complexity of traffic increased, policemen were employed in station areas to change the points at junctions and give instructions to drivers (Rowbotham 1999, p2). In 1838, the GWR issued instructions for hand signals to be given by the policemen. 'All right' was given when he knew the line to be clear ahead of the train, by holding his right arm out horizontally across the line of the rails that the train was running on. One arm held straight up above the policeman's head indicated 'caution', showing that the train should slow down and, stop was indicated by the policeman facing the oncoming train and holding both hands above his head (Hall(1) 2000, p17). This code was not universally applied. For example, on the LNE an arm held horizontally indicated danger, held at 45 degrees indicated caution and held upright indicated clear (Weightman 2000, pB2-1). However, consistency within an individual railway's operation was sufficient to ensure the clear understanding of signals at that time. At night, lights were used to give the signals. On GWR 'All right' was shown by a white light, 'caution' by a green light and 'stop' by a red light. Flags of the same colours were also used sometimes during the day (Hall(1) 2000, p17).

In order to improve the policeman's visibility to approaching drivers and enable him to signal from more than one location at a time, fixed signals soon began to appear. Initially these took the form of boards mounted on fixed posts alongside the track that could be pivoted to face end or side on and then left displaying that 'aspect' by the signalman whilst he operated other nearby signals. When the board could be seen face on, the driver was required to stop at it. When side on, he/she could proceed past. Signals of this form are still in use today on some minor lines in France, Germany and Switzerland. However, this basic design was rapidly developed into forms of signal that could display a positive indication of both stop and proceed, leading to the introduction of semaphore signals in 1841, based on a communication system used by the French Navy. Such signals are still in common use in the UK today, both on minor Network Rail lines and on the numerous preservation railways. In their modern form, semaphore signals display a red coloured arm, angled horizontally to indicate stop and at 45° to indicate proceed. As with the original policeman's arm signals, semaphore arms are supplemented at night by coloured lenses on the post end of the arm. These align with a lamp (initially oil, but electric in modern installations) in order to provide a coloured light indication at night (Hall(2) 2000, pB6/1; Rowbotham 1999, p2).

The policeman was not in a position to know whether the whole route ahead was clear, and could therefore only provide an indication that it was safe for the train to pass the discrete point at which he was located. From there the driver was still required to drive 'on-sight'.

When a train stopped in the station, the policeman had to stop any other approaching train. In poor visibility, they were instructed to go back 300 yards along the line in order to do this and (from 1841) to reinforce their signals by use of detonators (Hall(1) 2000, p17).

The 'on-sight' approach to train control is still in use on UK railways today on many goods lines, during shunting movements and in times of signalling failure. In these circumstances, the role of the policeman is adopted either by a shunt signal or the signaller / a shunter.

3.3 TIME INTERVAL WORKING

It was soon realised that it was difficult for a train to stop within its own sighting distance (Alston et al. 1971, p24). In order to reduce the risks associated with on-sight working, the railway policemen were therefore issued with egg timers and trains were only given authority to proceed past a policeman's assigned location once a fixed time interval had elapsed (Rowbotham 1999, p 2).

Unfortunately, there was a serious flaw in the approach of time interval working, as the policeman still had no way of knowing whether the line ahead was really clear or the previous train had in fact broken down, as was frequently the case (Alston et al. 171, p24; Technology Primer [(1)] 1997, p163). This gave rise to a particularly dangerous situation where poor visibility, or a tunnel in the route, prevented the driver from seeing an obstruction. There were also economic problems with time-interval working, in that time slots had to be made longer than strictly necessary in order to accommodate the variable performance of trains and their staff, meaning that capacity could not be optimised (Schmid 1999, p2). However, further progress had to await the invention of electrical communication.

Time interval working is not common now, but is still in use on some lines in the USA, where following trains in the same direction are authorised through a train order system (see section 3.6) to follow each other into an unsignalled single line based upon a time interval separation (usually 5 or 10 minutes). When this system of working is used, any train failing to keep to the normal speed of the section is required to drop a flare that has a five minute burn time. A following train is not then allowed to pass the flare whilst it is still alight. Similarly, if a train stops, the conductor is required to walk back a safe braking distance, place detonators and show a red flag or light to following trains (Connor 1999; Pachl 2000, p3/3-4).

3.4 SIGNAL BOX CONTROL

With the introduction of fixed signals, it soon became apparent that some form of co-ordination was required in place of individual railway policemen. This took the form of mechanical control of several signals and sets of points from a centrally located signal box (Rowbotham 1999, p3). In the UK, this was achieved using cables for the signals and rodding for the points, whilst cables were used for points as well in much of Europe.

The development of signal box control was mainly carried out as an economy measure, it being possible to operate all points and signals controlled from a signal box with only one policeman, who became known as the signalman or signaller (Hall[(1)] 2000, p25).

3.5 INTERLOCKING

With the signal box in place, the next major need was for a means of preventing the signalman from setting conflicting routes. By 1860 the first trials of mechanical interlockings had been carried out. These saw all of the levers to operate points and signals from a signal box mounted in a frame. When one of the levers was pulled, metal bars dropped into place to lock other levers, preventing them from being moved to allow a conflicting movement (Technology Primer [1] 1997, p163).

As time and technology progressed, the interlocking was supplemented by additional electric locking that prevented levers being moved when track circuits (see section 3.8) detected a train within the route. This was known as electromechanical interlocking (Howker[1] 1996 p B3/1).

The next significant change in the controlling medium was the introduction of relay interlocking in the 1930s. This saw the mechanical interlocking levers replaced by switches, with one switch per signalled route. A relay circuit was then designed for each route so that, when the signalman turned a switch to set up a route, the circuit would only be made if all track circuits in the route were clear, no conflicting routes had already been set and all points in the route were set and locked in the required positions (Kichenside et al. 1998, p145). Over time, this approach was developed to include an entrance and exit switch, operated in sequence to set up a route request (an approach refereed to as NX-operation) (Pachl 2000, p4/17). This form of interlocking is now commonly in use throughout the world.

By the 1950s, concerns over the design costs and time scales of new interlockings led to the development, initially in continental Europe, of geographical interlockings. These utilised packages of hardware that included everything that was needed for a signal, a single end of points or for double-ended points. The units were then laid out in the same geographical layout (as far as was possible) as the actual track layout and connected together with standard cables to complete circuits for route setting and signal aspect control. Further standard cables were also used for connection to the control panel and to the trackside equipment. In theory, this approach required minimal design, installation and testing effort and, whilst in practice the complexities of layouts often required additional elements of free wiring (typically accounting for 5 to 15% of the equipment on relay racks), the anticipated benefits were largely achieved. However, this was at the cost of a lot of redundant hardware within the geographical units (See Appendix A, Interview with Tony Rowbotham, question 12; Codd 1958, pp2,18; Pachl 2000, p4/18).

In the late 1970s various railways began looking at ways of using computers to handle the interlocking function, in order to reduce their size and maintenance requirements. The result in the UK was the British Rail Solid State Interlocking (SSI), which has subsequently been followed by a number of proprietary systems (Technology Primer [1] 1997, pp164-165). The introduction of computer based interlocking has seen major reductions in the capital and operating costs of large signalling schemes. "SSI's generate multiple cost savings. First the signal-box no longer has to accommodate a large relay room and can thus be smaller and cheaper. Second, manufacture of the previous electro-mechanical relays was labour intensive. Finally, relays have a finite life and require constant attention and replacement, where SSI is highly reliable and can be serviced by a single technician using replaceable electronics and modules" (Harris et al. 1992, p118).

As the overall cost associated with SSI type systems was less than that for relay based geographical interlockings, their introduction led to the end of the use of geographical interlocking within the UK. However, the attraction of geographical interlocking remained for railways with simpler, more standard layouts. They are therefore still used, along with relay interlockings, in other countries, such as South Africa (See Appendix A, Interview with Tony Rowbotham, question 12). A recent come back has also been made in the form of the Siemens

'SIMIS-W' interlocking, which implements geographic principles within a solid state interlocking (avoiding the need for redundant hardware by incorporating the geographic functions within the interlocking's software processes).

3.6 THE 'TRAIN-ORDER' SYSTEM

Ten years after the first interlocking came into use, the invention of the electric telegraph saw two distinct philosophies develop to replace time interval working. In North America the 'train-order' system saw a centrally located train dispatcher responsible for ensuring that all trains were in different places at any given time. In order to do this, the train dispatcher wrote the instructions to be followed in a book provided for the purpose. He then passed this order on to an agent for onward transmission by electric telegraph to the relevant local station (possibly via other agents). Each time the message was passed on it was also repeated back to confirm its accuracy. At the local station the instructions were written down for delivery to whoever was required to act on them. These individuals were then in turn required to sign for receipt of the instruction, and an acknowledgement that the instruction had been received was returned to the train dispatcher (Barwell 1983, p33). This kind of structured and recorded process ensured that action based on the dispatchers instructions would involve no risk of conflicting movements – assuming that the human operators all carried out their instructions correctly.

This type of system is still in use in Australia and the USA, where thousands of miles of remote railway are unsignalled, with trains being allowed to pass from one area to another by the use of train orders passed by radio between the train dispatcher and driver (Brotzman 2000; Pachl 2000, p3/3; Symons 2002, p1).

There are several modern equivalents of this system in Britain, France and elsewhere. One that is in every day use in Britain is Radio Electronic Token Block (RETB), which is used to control single line sections in rural areas. Each train operating under RETB is equipped with a speech and data radio unit with a unique identity. The signalman also has an SSI based system that allocates coded tokens to each section of track and prevents more than one token from being issued for a section. If the line is clear, the signalman issues a coded electronic token data message to the driver via the radio units. This requires simultaneous operation of buttons on both the signaller and driver's equipment. Once this has been done, the token appears on the driver's console display (in the form of the geographical name of the section that it is for). The driver then calls the control centre to confirm receipt of the token, before proceeding into the single line section. At the end of the section, the driver will see a clearance marker board. He then calls the control centre again to advise them that he is clear of the section and also uses the train's data radio unit to send the coded electronic token back to the control centre. Once again, this requires simultaneous activation of buttons on the signaller and driver's equipment, preventing the signalman from taking back the token without agreement, and the driver from returning it accidentally (Connor 1999; Hall 1987, pp103-4; Hall(1) 2000, p156).

The Train Order System is also used in the UK for single line working with staff or ticket, where the issuing of the staff or ticket to a driver acts as his/her authority, or 'train-order', to proceed into the single line. This method of working is still common on UK main lines during engineering work, when a pilotman, with signaller authority, issues a written ticket of authority known as a single line working ticket. This allows the driver to traverse a section of single line. It is usual practice for the pilotman to actually travel with the train, although this is not always the case. Where several trains are due to be sent through the section in the same direction, the pilotman must be personally present to instruct each driver, and is therefore permitted to send a train on and travel with a subsequent one. However, even when he travels with the train, the pilotman must brief the driver and then issue him/her with a ticket of authority. The detailed rules associated with this form of working can be found in Modules P (for pilotman working) and TS and TW (for staff working) of the rule book (Railway Safety, 2004)

3.7 BLOCK WORKING

The second philosophy developed as a result of the invention of the electric telegraph was the 'block' system. This operated on a much more localised basis through the use of special-purpose instruments, called block instruments, that used electric telegraph communication to control bell and needle indicators.

At the start of each block section there would be a signal box with a stop signal (box A). Initially, blocks were operated on the basis that the line was normally 'open' for traffic and the signalman was required to 'block' the line once a train entered it, not clearing it again until the train had passed the next signal box (box B). This meant that the signal at the entrance to each section normally showed a proceed aspect. However, an accident occurred at Abbots Ripton in 1876 because a signal froze in its normally clear position and could not be replaced to danger when a train had passed. Following that incident, the philosophy of block working was changed so that the line was considered normally closed, with the entrance signal held normally at danger (Weightman 2000, pB2-2). The signalman in box A was then only authorised to clear the entrance signal if the signalman in box B had accepted a train by use of the block telegraph instruments.

As a train approached box A, the signalman used the block instrument to send a bell code signal to the signalman at box B to ask whether the section of line between the two boxes was clear. If the previous train had left the block section, the signalman at box B would return a bell code for line clear, and turn the switch on the block instrument to 'line clear'. This automatically set the needle on the block instrument in box A to 'line clear', allowing the signalman there to clear his/her stop signal.

As the train proceeded past the cleared signal, the box A signalman would send another bell code for 'train entering section'. The signalman in box B would then set his/her block instrument to 'train in section', which automatically set the needle on the block instrument in box A to the same setting. The signalman in Box A then had to set his/her stop signal back to danger to protect the block section.

This procedure had to be repeated for each block section on the line. It embodied a great deal of redundancy, and was therefore inherently safer than the train-order system (Barwell 1983, p34). However, it placed the signalman under a lot of stress, particularly on a busy section of multiple track line. The signalman was required to visually inspect each train leaving the section in order to ensure that it was complete, by observing the red tail lamp on the last vehicle. (Corrie 1996, pC5/2). The signalman also had to remember which of the sections was occupied and, under the distraction of other movements occurring, could all too easily believe that a train had left a section when it had not, and ring back 'line clear' in error (Goddard 1997, pA3-1).

The adoption of block working was initially unpopular with some railway authorities, who saw it as impeding the flow of traffic on the line (Weightman 2000, pB2-2). It was also expensive, leading other companies to delay its implementation either due to lack of funding or concern for shareholder interests. The use of time interval working was therefore perpetuated until a number of serious accidents resulted in the use of absolute block working being made compulsory for all passenger lines in the UK by the 1889 Regulation of Railways Act (Hall[1] 2000, p55).

Some variants of the block working system were implemented in various parts of the world to introduce further automation, and reduce dependency on the signalman. One such system was installed on the Glasgow Cable Subway. It employed treadles at station exits that controlled an electric bolt interlock attached to the starting signal lever mechanism. The electric bolt was only released once a second treadle at the exit of the block section had been operated, thus automatically preventing the starting signal from being cleared with a train in the block section ahead. (Barwell 1983, p34)

In general, however, the problem of knowing when a train was in section was solved by the introduction of track circuits.

3.8 TRACK CIRCUITS AND TRACK CIRCUIT BLOCK

The use of track rails to act as electrical conductors for signal purposes was first suggested in mid-19th century British patents, but it was not until 1872 that the 'steady-energy track circuit' was first used, in America (Duckitt 1967, p1). The track circuit consisted of a voltage source connected across the running rails at the beginning of a section of track and a relay connected between the rails at the other end of the section. The relay would therefore be energised when no train was present. As a train occupied the track section, the train's axles short circuited the track circuit causing the relay to de-energise, indicating the presence of a train. This system had inherent safety properties, in that a failure of the power supply, accidental disconnection of a connecting lead from the rail and the most likely failure conditions in the relay would all cause the circuit to 'fail safe' by indicating the presence of a train even if there was not one there (Corrie 1996, ppC5/2-3).

The invention of the track circuit enabled the block system to be enhanced. A track circuit at the entrance to each block section provided a physical indication of the presence of a train entering the section and automatically registered 'train in section' on the block instruments. In addition to this, an output from a track circuit could be used to energise an electric lever lock on the mechanical interlocking when the presence of a train was detected, thus preventing the signal being changed to line clear (Technology Primer [1] 1997, p163). Track circuits also provided a means of indicating the presence of a train stopped in front of a signal to the signalman, to act as a reminder to him/her that the train was there (Rowbotham 1999, p5).

In time more and more track circuits were installed, which led to the development in the 1920s of 'track circuit block' signalling. This system saw the section between signal boxes divided into several track circuits. Automatic signals were then introduced, worked solely by the operation of the track circuits, no action by the signaller being necessary (Rowbotham 1999, p5). This type of system is still in use today in USA, where it is known as 'Automatic Block' Signalling (IRSE [2] 2001, p18) and in the UK for automatic signalling sections.

By this stage in the development of signalling systems, the control system of Figure 3-2 had been developed into the more advanced control system represented in Figure 3-3. The signalling system had now begun to close the control loop automatically, as had the trainborne system's provision of speed information on a speedometer. However, observation of signal aspect and the feedback loop determining the error that exists between that target and the train's actual movement were still provided by the driver, not the system.

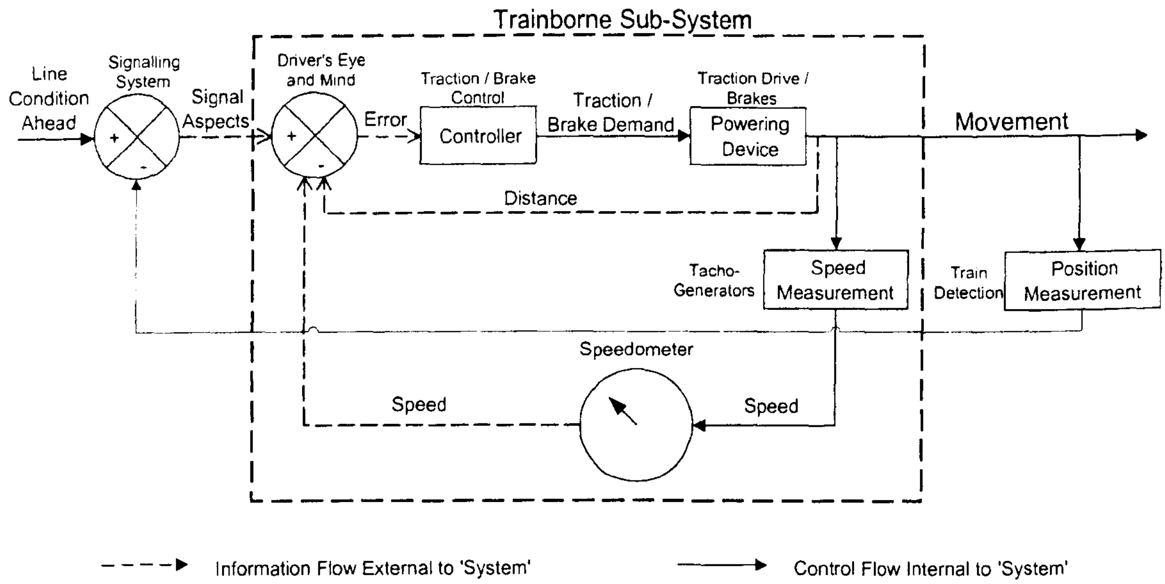


Figure 3-3: Advanced Human-Machine Train Control System (see Appendix B)

As railways continued to progress, the need for increased safety saw the introduction of enhancements to both the signalling system and the information flow between the signalling and the driver. Most of these changes made no material difference to the control system shown in Figure 3-3. They are discussed here, however, in order to explain the control systems that are currently in use.

3.9 COLOUR LIGHT SIGNALLING

The next major development in signalling was the introduction of colour light signals. The first UK installation was in 1924, on the line out of Marylebone station. The new electric lamps were much brighter than the old oil lamps that had been used to provide visibility of semaphore signals at night. They also had the advantage of providing the same indication to drivers by day and by night (Rowbotham 1999, P6).

By the use of electric colour light signals, the location of signals was no longer constrained by the strength of the signalman. Instead, the separation of signals began to be determined by a combination of line speed and braking performance (Technology Primer [(1)] 1997, p165).

3.10 ROUTE AND SPEED SIGNALLING

At this point, it is worth noting a divergence that occurred between UK influenced signalling and the signalling practices used in other parts of the world.

Railways in the UK developed a system of route signalling, where the signal aspect indicated the route that had been set and the driver then had to use his/her own route knowledge to determine the authorised speed. However, in those countries where British influence was not so strong, including most of Europe and North America, a system of speed signalling became the norm. Under this type of system the signal aspect authorised the driver to proceed at a given speed (Knowles et al 1999, p16).

Speed signalling was originally developed from semaphore arms, and at that time its main advantage was one of economics – with savings in arms, lights and masts being offered, together with significant reductions in wear and tear and increases in track capacity (Crook 1993, pp145-6). However, with the- introduction of colour light signalling, most of the benefits attributed to speed signalling were also realised in the route signalling approach. In contrast, attempts to convert speed signalling semaphore signals to colour light signals led to the requirement for a plethora of aspects, in effect making speed signalling less economic than route signalling. The variety of aspects, each of which represented a target speed to the driver, also became far more complex than the UK route signalling system, creating more scope for driver error. As a result of this, a technical paper to the IRSE in 1982 was able to state in considering the drawbacks of colour light speed signalling systems, "thank goodness we in the UK avoided (eventually) that pitfall" (Wyatt 1982, p70 and 71).

Despite this strongly expressed view, subsequent discussions reported within the same paper presented a more positive view of speed signalling. In particular, it was noted that the advent of Cab Signalling and Automatic Train Control (see section 4.3) had introduced a requirement for signalling information related specifically to speed, whilst also providing a means of relaying speed information to the driver without any real limit on the number of 'aspects' that can be handled (Wyatt 1982, p86). This development offers the opportunity to benefit from the advantages of speed signalling, without the main historical drawbacks.

It has been recognised that the UK route signalling practice inherently requires drivers to possess route knowledge and interpret the meaning of signals, and thus creates a system dependence on driver infallibility. The IRSE's Signalling Philosophy Review in 2001 noted "humans remain less predictable than control systems and will always have the occasional 'bad day'." The review therefore concluded that there was a basic choice for UK signalling – to individually mitigate the driver issues that stand to undermine UK route signalling practice, or to introduce ATP (IRSE(1) 2001 section 62). This view was also echoed by the IRSE International Technical Committee in a supporting document, where they concluded that "a positive outcome of the introduction of ATP systems in the UK will be to take the interpretation of information and the supervision of safety related actions into the system, reducing the dependence on route knowledge and reducing the opportunities for human errors to affect safety" (IRSE(2) 2001 p2).

At the same time, practitioners of route signalling practice argue that the flexibility for drivers to use their judgement in interpreting signal aspects in the light of their route knowledge means that the route signalling practice can provide better use of capacity and better regulation of traffic than speed signalling (Taskin et al. 1995, pA3/3). It would therefore appear likely that the practice of route signalling will be maintained in the UK in the short term. However, it is also likely, and certainly desirable, that UK signalling will in time adopt ATP and subsequently in-cab speed signalling (and hence a speed signalling approach) to replace existing route signalling systems. This has already begun to happen on LUL's Central Line, where the particular suitability of Speed Signalling to automation has seen the use of in-cab speed signalling within ATP protected manually driven modes, based on the information already required by the system for operation in the fully automatic mode.

3.11 FIXED BLOCK COLOUR LIGHT SIGNALLING

Historically, most signalling schemes have been designed to fixed block principles, in which the track is divided into a series of fixed sections (called blocks), as discussed in sections 3.7 and 3.8. The operation of such schemes is based on two main principles:

1. a train can not be authorised to enter a block if there is already a train occupying it and;
2. the distance separating two trains must always be greater than the braking distance required for the rear train to stop.

   (Taskin et al. 1995, ppA3/2-3)

With the advent of the colour light (discussed in section 3.9), the fixed block approach was developed into a range of systems that are now widely used throughout the world, known collectively as fixed block colour light signalling. Under these systems, the fixed block principles are maintained by controlling the aspect sequence behind a train. This can be done by use of either route or speed signalling, the difference in application being the meaning of the aspects to the driver. However, for the purposes of clarity the subsequent discussions within this section will be based on the UK route signalling approach.

3.11.1 TERMINOLOGY OF FIXED BLOCK SIGNALLING

Before continuing with the discussion of fixed block signalling, a few definitions of terminology may assist the reader:

- **Route:** "a predetermined path for a traffic movement" (BSI 1998, p8);
- **Before; In Rear:** "anything on the approach side of a given point when facing the direction of travel" (Jackson 1997, p236);
- **After; Beyond; In Advance:** anything beyond a given point when facing the direction of travel (Jackson 1997, p236);
- **Junction protection signals:** Any signals displaying cautionary aspects on approach to a junction are termed junction protection signals. The last such signal on approach to a junction is referred to as 'the' **junction protection signal** (or sometimes simply as the junction signal);
- **Home signals:** Any signals displaying cautionary aspects on approach to an occupied platform are termed home signals. Where there is more than one such signal on approach to a platform, the first encountered is often referred to as 'the' home signal or the 'outer home'. The innermost home signal (that closest to the platform) is generally referred to as the 'inner' home and any intermediate signals as 'intermediate' home(s). In a main line 3 or 4-aspect system, only the inner home would display a red aspect with a train occupying the platform. However in LUL 2-aspect signalling, all of the home signals would display red aspects with a train in the platform, clearing to proceed aspects in sequence (starting with the outer home signal) as the train pulls away.

3.11.2 2 ASPECT SIGNALLING

The most basic form of colour light fixed block signalling is the 2-aspect system. The use of this approach originated on LUL, where speeds were relatively slow, braking rates high and sighting conditions close to ideal in most places. Under this system a stop signal at the entrance to each block section displays a red aspect if there is a train in the block ahead. Drivers are then required to stop their train before this signal. With no train in the block the signal displays a green aspect, authorising the driver of an approaching train to proceed beyond it.

Due to the fact that driver error and system failures can cause a train to proceed beyond a stop aspect without authority, an additional safety feature is usually included within the block signalling approach. This is idea of an overlap, or distance beyond the exit to a block section that must also be unoccupied by trains before the signal at the block's entrance is allowed to give a proceed aspect. By LUL's standards for signalling, the overlap is required to be a calculated, worst case, stopping distance under emergency braking (an approach which supports the use of a trainstop based ATP system – see Appendix C). On UK main line railways, it is usually assigned a nominal length of 180m (Marks 2000, p5).

Even where stopping within the overlap is enforced, it is desirable for both passenger comfort and service reliability that the train should be stopped before the signal. This means that the driver needs to be able to see the signal sufficiently before reaching it to be able to react to the observation of a red aspect and stop under service braking rates before passing it. The point at which the signal must be clearly visible to ensure that the driver has time to observe it is known as the 'sighting point', and the distance between this point and the location at which braking must commence to stop by the red aspect is known as the 'sighting distance'.

If the driver would be unable to observe the signal from the sighting point (in all weather conditions), a yellow/green repeater signal (or more than one if required) can be located in between. This signal displays a green aspect if the main signal is green, and a yellow (caution) aspect if the main signal is red.

Based upon this information, it is possible to represent a two aspect signalling system, as shown in Figure 3-4. For the purpose of simplicity, this has been done with reference to headway based upon train following capacity (see section 2.5 for definitions of these terms).

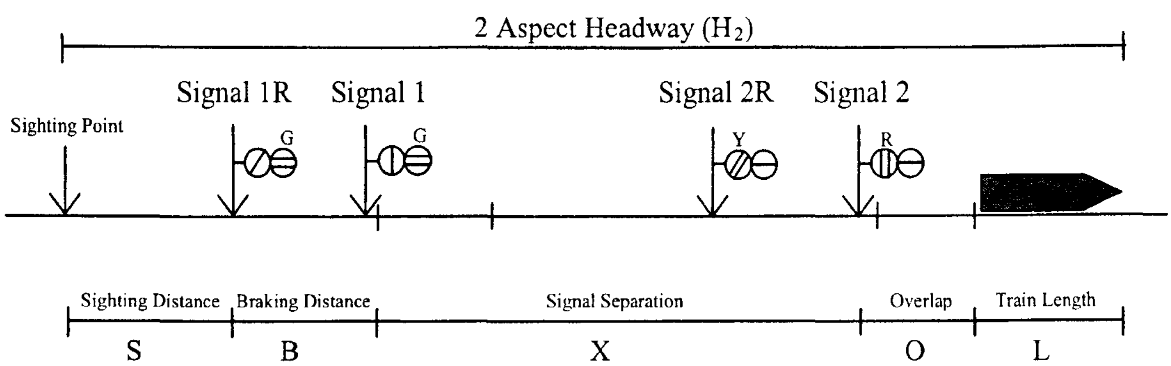


Figure 3-4: 2-Aspect Fixed Block Signalling Layout

From Figure 3-4, it can be seen that a 2-aspect fixed block signalling system has a headway distance given by:

$$H_2 = B + X + S + O + L \qquad \text{Equation 3-1}$$

Which means that in order to achieve a specified headway, the greatest permissible distance between consecutive stop signals is given by:

$$X = H_2 - (B + S + O + L) \qquad \text{Equation 3-2}$$

In theory, there is no minimum distance by which 2-aspect signals must be separated. This is particularly true if speeds are low enough and the track straight enough to allow clear sighting of signals within the required braking distance without the need for repeaters. In practice, however, there are several constraints that will apply:

- The closer the signals are, the more signals will be required and thus the more expensive the system will become to provide;

- As speeds and braking distances increase, it becomes necessary to provide repeaters for all signals if sufficient sighting distance it to be provided. This further increases the system costs;
- In practice, there will actually be a limit determined by the combination of speeds and signal separations for which X becomes equal to B. When this occurs, the repeater for one signal has to be located on the same post as the previous signal. At this point, the green aspect of the repeater effectively becomes superfluous;
- As the location of a required repeater signal becomes close to that of the previous signal, the combination of aspects displayed could lead to driver confusion. If the separation were to be allowed to become smaller than the braking distance, the repeater would need to be located before the previous signal, making the driver's task virtually impossible.

3.11.3 3 ASPECT SIGNALLING

An alternative approach that is utilised to overcome some of these limitations is 3-aspect signalling. Under this approach, each signal is capable of displaying a red (danger), yellow (caution) and green (Proceed) aspect, with the sequence operating as shown in Figure 3-5.

A 3-aspect fixed block signalling system with signal spacing 'd' has a headway distance of:

$$H_3 = 2d + S + O + L \qquad \text{Equation 3-3}$$

Giving the greatest permissible distance between consecutive stop signals in order to achieve a specified headway to be:

$$d = \frac{H_3 - (S + O + L)}{2} \qquad \text{Equation 3-4}$$

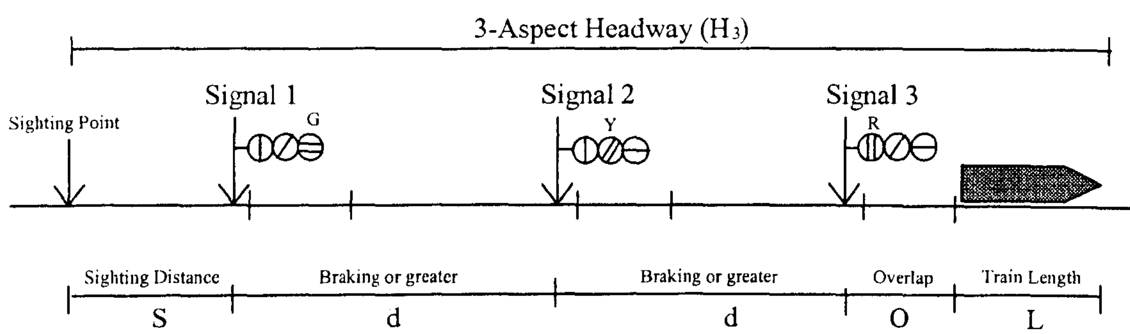


Figure 3-5: 3-Aspect Fixed Block Signalling Layout

Consideration of the equations for two aspect-signalling reveals that the headway time achieved when the signal separation is equal to the braking distance is the same as that for a 3-aspect system. However, the 2-aspect approach under these conditions would require four aspects on each signal post (the green and red of the main signal and the yellow and green of the next signals repeater), whilst the 3-aspect approach only requires three. The 3-aspect approach therefore has the advantage of being more economical than the 2-aspect approach in enabling trains to operate at this headway time.

The scale of this economy also means that (assuming all signals require a repeater) "2 aspect signals with a section length of as little as twice the braking distance require the same number of signals as 3 aspect signals" (Catalis Rail Training[2] 1999, p12). This means that when signal separation is required to be between 1 and 2 times the braking distance, 3-aspect signalling effectively provides a means to improve capacity over 2 aspect signalling at no extra cost.

In practice, 3-aspect signal separation is usually kept between 1 and 1.25 times the braking distance, extended to 1.5 times in exceptional circumstances (Weightman 2000, pB2-7).

3.11.4 4 ASPECT SIGNALLING

The best achievable headway under the 3 aspect system will be when 'd' is equal to the calculated braking distance, which is not actually an improvement on that achievable by use of 2 aspects. If a better headway is required than this, then a means needs to be found of bringing the signals closer together than braking distance, whilst still providing the driver with a warning that he is approaching a red aspect in sufficient time for him/her to be able to stop at it. This can be done by use of a 4 aspect system, where the red (danger), yellow (caution) and green (Proceed) aspects are supplemented by a double yellow (preliminary caution) aspect, with the sequence operating as shown in Figure 3-6.

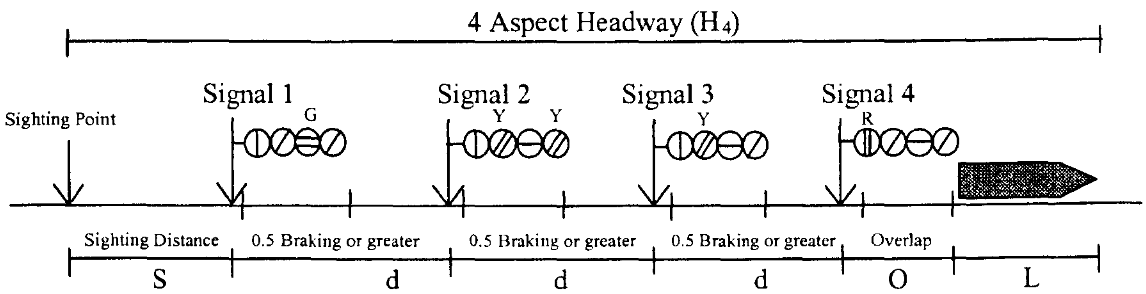


Figure 3-6: 4-Aspect Fixed Block Signalling Layout

A 4-aspect fixed block signalling system with signal spacing 'd' has a headway distance of:

$$H_4 = 3d + S + O + L \qquad \text{Equation 3-5}$$

Giving the greatest permissible distance between consecutive stop signals in order to achieve a specified headway to be:

$$d = \frac{H_4 - (S + O + L)}{3} \qquad \text{Equation 3-6}$$

The best possible headway under this arrangement will be when the average signal spacing is equal to half the calculated braking distance. Under this condition, the headway achieved is an improvement on both the 2 and 3 aspect systems. A further advantage over the 3-aspect system is that "in practice, 4 aspect signals do not have to be equally spaced, as long as the distance from the double yellow to the red in each case is not less than braking distance." (Catalis Rail Training[(2)] 1999, p8). However, this approach does require more signal aspects to be installed than either the 2 or 3 aspect approach.

In practice, signal separation in a 4-aspect system is usually arranged so that there is between 1 and 1.25 times the braking distance between the first caution aspect (double yellow) and the stop aspect (red). As with 3-aspect signalling, this may be increased to 1.5 times the braking distance in exceptional circumstances. In addition to this, the distance between the second caution (yellow) and stop aspect must be at least one third of the distance between first caution and stop aspect (Weightman 2000, pB2-7). These 'rules' are generally applied in order to avoid driver confusion by irregularly spacing signals, the aspects of which represent an instruction to stop at a signal ahead.

3.11.5 5 OR MORE ASPECT SIGNALLING

A similar methodology could in theory be adopted for systems with any number of aspects. Simply dividing the braking distance into additional sections gives the general headway distance equation:

$$H_n = (n-1)d + S + O + L \qquad \text{(where n>2)} \qquad \text{Equation 3-7}$$

Giving the greatest permissible distance between consecutive stop signals in order to achieve a specified headway to be:

$$d = \frac{H_n - (S + O + L)}{n-1} \qquad \text{Equation 3-8}$$

For a required braking distance (B), the best possible headway for an 'n' aspect arrangement (where n>2) will therefore be when the average signal spacing (d) is equal to:

$$d = \frac{B}{n-2} \qquad \text{Equation 3-9}$$

This basic approach can be adopted for the addition of any number of aspects. However, as demonstrated by Figure 3-7 and Figure 3-8, the implementation of signalling systems would progressively become more and more expensive for proportionately less benefit each time an additional aspect was added. Both figures assume typical UK main line operation at a constant velocity of 60mph with a 275m sighting distance, overlap of 180m and train length of 200m.

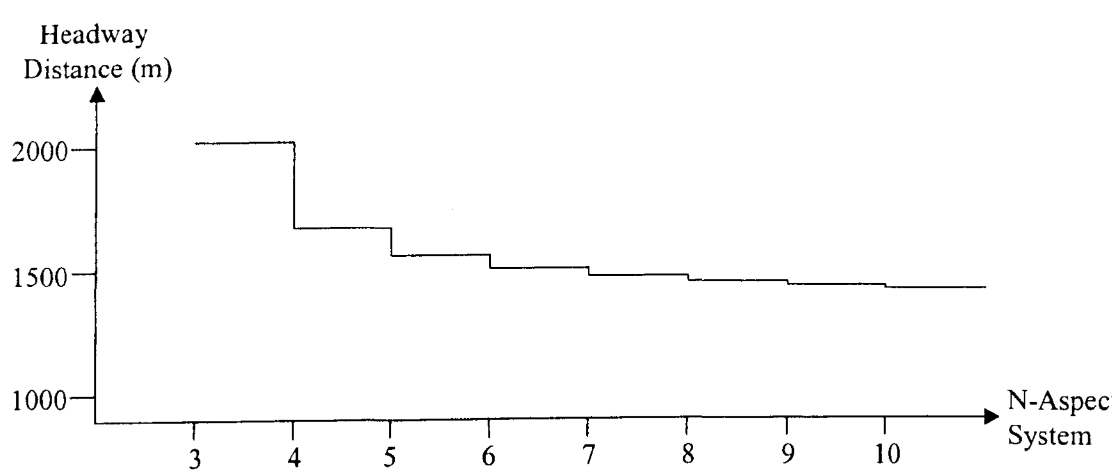


Figure 3-7: n-Aspect Fixed Block Signalling Headway Distance

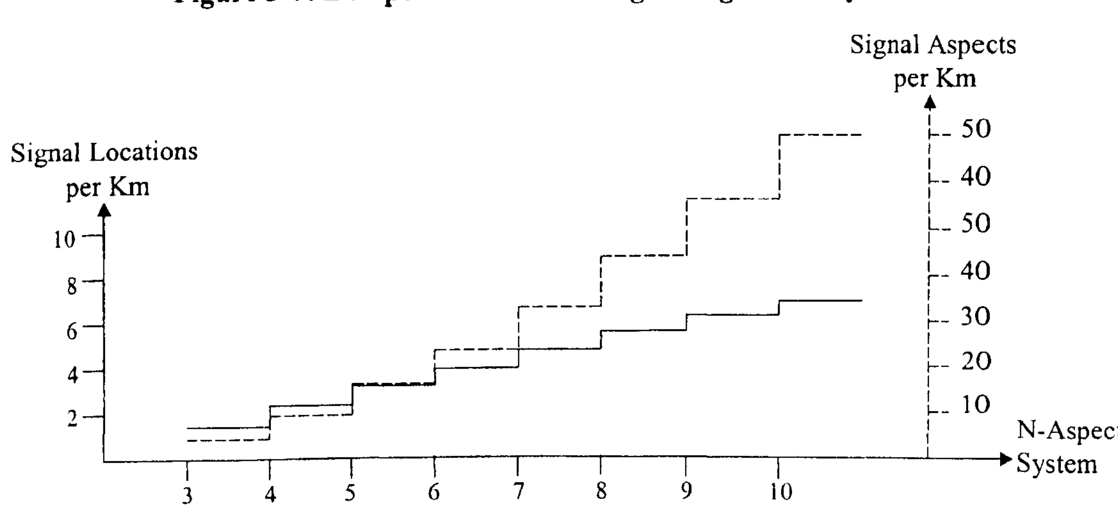


Figure 3-8: n-Aspect Fixed Block Signalling Infrastructure

The additional expense associated with having more aspects has in part been overcome for application in critical areas in the UK by use of flashing aspects. Flashing aspects are effectively a fifth aspect without needing to provide additional heads, just additional aspect controls. The use of a flashing double yellow aspect, followed by flashing yellow and then steady yellow aspects, is now common practice to permit high-speed divergence at junctions.

With the introduction of 140mph test running on the East Coast Main Line in 1991, the existing 4-aspect signalling system could not provide sufficient braking distance from the first caution aspect to a red signal. To move signals in order to provide sufficient braking would have been prohibitively expensive and caused unacceptable reductions in line capacity. BR therefore decided to introduce a fifth (flashing green) aspect rather than increase signal separations (Kitchenside et al. 1998, p184-5). The use of this practice was not extended beyond trials.

Besides the expense of providing a system of lineside signals with 5 or more aspects, the complexity imposed upon the driver of interpreting large numbers of aspects must also be considered. "It is generally used that it is impossible to go beyond four aspects otherwise the driver's task becomes too complicated" (Weber 1975, p2).

3.11.6 HEADWAY LIMITATIONS

It can be seen from the headway equations of all fixed block approaches that that the majority of factors (B, L and possibly O) are specific to the characteristics and performance of a given train, and that most factors (S, B and possibly O) are also dependent upon the speed of operation. This means that the achievable headway of a traditional fixed-block signalling system must be designed in, accounting for the trains performance characteristics, permitted speeds and the operational requirements of the railway. Thus the system is fairly inflexible if the requirements for headway, train performance or line speeds are changed.

It is also worth noting that in all of the fixed block signalling approaches illustrated in sections 3.11.2 to 3.11.5 the best achievable headway distance (when all trains are following each other at the designed line speed) exceeded the calculated braking distance by at least the combined length of the train and the provided overlap and sighting distances. For a realistic application, the headway distances would actually be significantly higher than the best-case conditions.

Where the line is utilised by homogenous trains, the position and length of block sections can be optimised for the particular performance of those trains, but that optimisation must make assumptions concerning the actual train and driver performance. If different stock types are to use the section of line, fixed block signalling must be designed to provide safety (sufficient sighting, braking and overlap distances) for the 'worst case' stock characteristics, with the result that the designed signalling will not allow optimal use to be made of other stock types.

Assuming that the headway distance is travelled at a fixed velocity, V, the equations for headway distance already considered can be modified to provide a headway time (the time interval between trains supported by the headway distance):

$$Ht_2 = \frac{B + X + S + O + L}{V} \qquad \text{Equation 3-10}$$

$$Ht_n = \frac{(n-1)d + S + O + L}{V} \qquad \text{(where n>2)} \qquad \text{Equation 3-11}$$

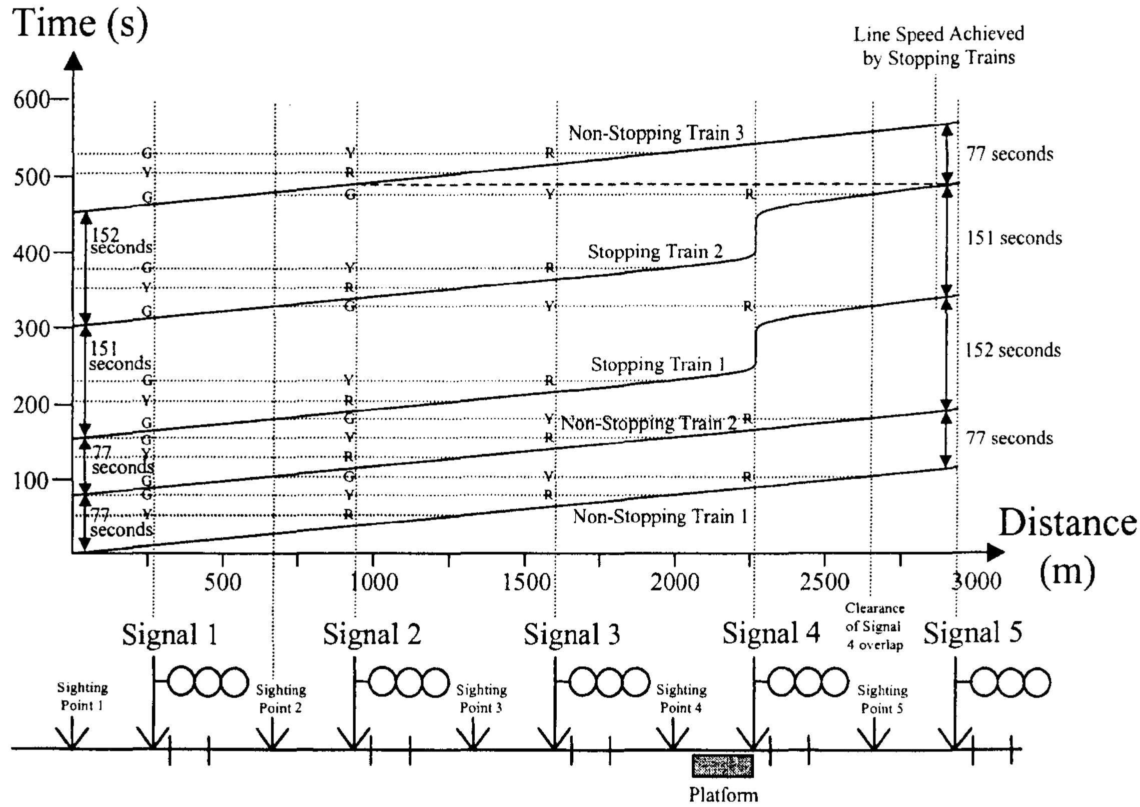
If these equations are plotted for varying speeds, it can be seen that increasing speed alone does not necessarily provide a means of improving the system headway (see 'train following headways' in Figure 2.12). Clearly the line speed therefore plays a significant factor in the determination of headway, and in most cases there will be a trade off required between headway and journey time.

The consideration of headway as a time between trains becomes particularly relevant when it is considered that the headway impact of fixed block systems is not only experienced whilst trains are following each other at the designed line speed. When trains are operating at lower speeds, the braking distances needed are also reduced, and the separation between trains required for safety can therefore become significantly less. However, the fixed nature of the block sections and their associated signals and track circuits will not permit this, but enforce continued separation by more than the line speed braking distance. Under these conditions, the designed headway distance may well still be maintained, but the headway time between trains will become significantly higher. As a result, if trains operate at lower speeds than the design speed the designed headway distances will effectively be enforced by the system, but it will not be possible to achieve in practice anything like the designed headway time (Pope 1975, p21). Consideration of headway time therefore presents a much more useful and powerful tool than consideration of headway distance.

In any practical railway system trains must stop at stations to allow passengers to board and alight. In main line railway operations it is generally assumed that non-stop trains will run through stations at full speed, whilst other trains will be required to stop at the same station. The signalling system must therefore be designed to support the safe operation at full speed, but the capacity built in to such a design will in effect be wasted by the stopping service. This can be seen clearly from the graphs in Figure 2-9 and Figure 3-9 (which, for the conditions modelled, shows that stopping trains require approximately twice the headway time of non-stopping trains). In practice, measures can be taken to minimise this effect, since:

a) The headway increases directly with the duration of the station dwell;
b) Headways can be reduced by placing signals at minimum separation on approach to stations;
c) Improvements to train performance (both braking and acceleration) will also enable improved headway performance;
d) Lowering line speeds around the station area will reduce braking distances and allow smaller signal separations, and subsequently improved headway (at the expense of lower journey times for non-stopping trains);
e) The following train can not approach the station at full speed until the first train has cleared the overlap. Placing the first signal after the station at the end of the platform will therefore minimise the effect of stopping trains on headway.

(Pope 1975, p25).



Design Assumptions:

Sighting distance = 275m
Train length = 200m
Braking rate = -0.54 m/s/s
Station Dwell time = 30 seconds

Overlap length = 180m
Line Speed = 60mph = 25.8 m/s/s
Acceleration rate = 0.56 m/s/s

Distances and Times determined by use of Newton's Laws Of Motion.

Figure 3-9: 3-Aspect Fixed Block Signalling Headway Times

Where speeds are sufficiently low to permit the use of 2-aspect signalling without repeaters, as is typically the case on metros such as LUL, the signal layout design can be optimised by placing signals at less than braking distance separation on approach to the critical station area. Such signals are generally referred to as 'home signals', and are located so that the overlap after the outermost home signal ends before the platform (allowing a following train to draw up as close to the station as possible whilst the first train is still in the platform). Additional home signals can then be added closer in to the station, allowing the following train to approach the station as the fist train leaves. By use of this technique, LUL generally anticipate that:

- One home signal enables a service of 38 tph (95 seconds between trains);
- Two home signals enable a service of 41.5 tph (87 seconds between trains);
- Three home signals enable a service of 43 tph (84 seconds between trains);
- Four home signals enable a service of 44 tph (82 seconds between trains).

  (LUL 1994, p6).

In all but the simplest railway layouts trains will also encounter junctions, most of which require lower speeds of operation for diverging routes than for the main route. Conditions such as these also act as constraints on the capacity achievable by the signalling system and require consideration during design. Further consideration will be given to this in Chapter 8.

3.12 SUMMARY

In this chapter the author has demonstrated the variety of signalling solutions developed over the years, and has provided some of the background reasons behind each development.

A number of common factors can be traced throughout the development of the various signalling systems that have been described in this chapter. One of the most significant of these is the fact that the operation of a railway, including its control functions, has traditionally been split between a number of departments that have achieved their purposes by utilising a number of technical systems:

- The operations functions (planning and day to day operation of the railway services, including station operation, train operation and signal control) have been performed by drivers, station staff and signallers / dispatchers, working for the operations department;
- Design, construction and maintenance of railway equipment has been divided between departments responsible for specific engineering functions, usually:
    - civil engineering (responsible for buildings, structures and the track formation);
    - mechanical & electrical engineering (responsible power supplies, lighting and lift / escalator equipment);
    - rolling stock engineering (responsible for the rolling stock and all trainborne control subsystems);
    - signalling & communications engineering (responsible for other control sub-systems).

This split in responsibilities has led to a disjointed approach to railway control and has, in particular, acted as a limiting factor in the development of train control systems. In chapter 4, the author will consider some more modern approaches that are beginning to break down these departmental / functional divisions in order to consider the overall control system as a complete entity.

4 ADVANCED METHODS OF TRAIN CONTROL

4.1 INTRODUCTION

One common factor in all of the signalling developments described in chapter 3 is reliance on human operators, in the form of drivers and signalmen. Most of the functions carried out by the driver are mundane and repetitive. His/her main function is to drive the train, controlling acceleration and braking in response to the signalling system, whilst observing everything on or about the line as the journey proceeds. In doing this he/she is required to consider adherence to the timetable, energy use and passenger comfort and, where necessary, to communicate with other staff on the train and with the controlling signal box. He/she also carries out train preparation, undertakes station duties and manages any fault alarms that may occur on the train (IRSE(4) 2001, p36 section 7.7).

In practice, drivers carry out these tasks with remarkable reliability. For example, the mean time between passing signals at danger has been estimated to be over 17 years for the average driver, an error rate that is considerably lower than that quoted in generic human reliability studies (IRSE(4) 2001, p52 section 11.2).

Similarly, signalmen are required to "control the safe movement of trains in accordance with the rules and regulations" (Francis 1994, p1) in conjunction with controllers who oversee the route and are responsible for regulating traffic to optimise capacity (State of the Art 1998, p341). "The signaller has two primary responsibilities: planning the movement of trains and ensuring that points and signals are set so that these movements can be completed in safety" (Fenner 2002, p29). Once again, these functions are performed extremely reliably. However, despite their high reliability, human being do make mistakes and, as a consequence of this, the root cause of most accidents is human error (Burrage 1996, pA3/2).

Even at such a low SPAD rates, in one hundred years of railway operation (from 1900 to 1999) the four major British rail routes (the West Coast, East Coast, Great Western and South Western Main Lines) experienced:

- 42 accidents following SPADs resulting from driver error;
- 22 accidents caused by driver error in overspeeding;
- 24 accidents as a result of signalman error.

In all, driver error accounted for over 44% of accidents on those routes and signalman errors for over 16% (giving a total of 60% of accidents that resulted from human error) (Xue 2002, p7).

In the last twenty years of this period (from 1980 to 1999), accidents due to signalman errors appear to have reduced significantly, but those due to driver errors have increased when compared with the average. The result has been:

- 9 accidents (37.5%) following SPADs resulting from driver error;
- 3 accidents (12.5%) caused by driver error in overspeeding;
- 1 accident (4.2%) as a result of signalman error.

In all, despite developments in train and signalling control systems, the number (13) and percentage (54%) of accidents that resulted from human error between 1980 and 1999 remained high, making driver error the most significant cause of accidents on the railway (Xue 2002, p7).

There are two approaches that can be taken to making technological work systems involving human operators (such as a railway control system) safer. The first of these is to assess how reliable individuals are at what they do, and to design the system accordingly. This involves "measures to reduce the probability of error, for example by improving the visibility of signals" (Short 2001, p1). A number of factors can be considered in doing this, including:

- Analysis of the individuals' characteristics (such as alertness, attention, motivation, skills / knowledge acquired, training, task contentment and physical condition);
- Analysis of the workload and consequent stress associated with a particular task;
- Analysis of the work environment and interface design (ergonomics);
- Analysis of the organisation (procedures/instructional design, organisation structure and management safety culture).

The combination of these factors are commonly referred to as 'Human Factors'. That is "the interplay between the operator, the machinery and working environment" (Cooksey 2001, p4)

It has been concluded by some human factors experts that "attention to 'soft solutions', such as management culture and organisation, training procedural design and application of ergonomics design principles give a far more cost-effective approach to risk reduction than that offered by traditional engineering solutions" (Smith 1999, p8). However, considering the already higher than expected reliability of railway operators, the 2001 public inquiry into train protection systems noted that "the view of the HSE is that we are already close to what can be expected of human response in the cab or signal box, a view to which we agree" (Uff et al. 2001, p114). This view was also reflected in an IRSE study into signalling philosophy and human factors, which found that "experience has shown that human error resulting in catastrophic accidents can occur even where the design of the signalling system and the training and management of drivers appear near to optimum from a human factors point of view." (IRSE[(1)]2001 Paragraph 109). It therefore appears that, whilst it is important to be aware of the human factors involved within the engineering design of railway control systems, consideration of human factors alone will not be enough to improve safety significantly.

The second approach that can be taken to making technological work systems safer is to accept that people vary in their performance and that, inevitably, they will make errors. This "leads us to look at implementation of systems such as ATP (Automatic Train Protection), where the aim is to prevent or mitigate the effects of an error" (Short 2001, p1). In line with this approach, the IRSE in their review of UK signalling philosophy concluded that, since human error is such a major cause of accidents, the way to ensure safety is to "circumscribe the role of humans with safety devices that will eliminate the consequences of human error" (IRSE[(1)]2001 Paragraph 49). In other words, the introduction of automation to the process of train control.

In recognition of the potential for significantly improving safety, the author primarily considers the subject of automation in train control in this chapter. However, automation is not the only advance from traditional signalling methods that has occurred in the field of train control. In particular there have been advances in technological ability that have led to the development of transmission based signalling systems. These in turn have opened up the potential for the replacement of fixed wayside signals with in-cab signals, for developments in the concepts of train separation and ultimately provided the required medium for more advanced forms of automation. The chapter therefore begins with a discussion of this development in transmission of signalling information.

4.2 TRANSMISSION BASED SIGNALLING

Traditional signalling systems have relied upon the transmission of audible or visual information between the trackside and the train in order to provide a train driver with the authority to proceed. As discussed in chapter 3, this was initially achieved through direct human communication, but later began to utilise 'hard wired' mechanical or electrical transmission of information within the signalling system. In the case of a 'train-order' system, this took the form of audible indications to a human agent (via a telegraph system), which could then be written down and physically passed to the driver. Similarly, under block working, the telegraph system was utilised to provide a combination of audible and visual information to a signalman, who then used mechanical transmission to issue visual indications (in the form of signals) to train crew. As signalling progressed to track circuit block, and subsequently to colour light signalling, the use of verbal and mechanical transmission gradually declined, in favour of visual and electrical transmission. By use of these transmission techniques, it was possible to develop increasingly complex wayside signalling systems. However, as long as the only medium available for transmission between wayside equipment and trains remained that of visual signals (and occasionally audible ones in the form of detonators), there was very little that could be done to develop the interface between the signalling system and the driver. This interface, heavily dependent on human factors such as the driver's attention and perception, therefore formed a major constraint to the further development of train control.

In order to overcome this transmission constraint, a number of systems were developed to enable communication between trainborne and wayside equipment. These included purely mechanical systems, (such as trackside treadles, trainstops and trainborne plungers), electro-mechanical systems (designed to make a mechanical contact over which an electrical current could pass to complete circuits on-board the train) and later systems utilising electro-magnetic coupling to achieve the same functionality without the need for actual physical contact. These types of system were traditionally developed to augment an underlying signalling system. By their use, it became possible to provide simple audible and visual indications to the driver from within his/her cab, and also to enforce automatic operation of the train's braking systems. As the technology by which this could be done progressed, so the constraints placed on system functionality by that technology gradually reduced. Detailed descriptions of a number of systems developed to use these transmission principles are given in Appendix C.

Considering the number and variety of transmission mediums utilised over the years within the context of signalling and train control, it is perhaps surprising to find the contrary view expressed that 'transmission based signalling' represents a relatively new advancement on the traditional signalling system. The explanation for this comes largely from the fact that we have only recently seen the development of secure mechanisms for the transmission of data without fixed connections, combined with the advent of sophisticated and reliable systems for the processing of that data.

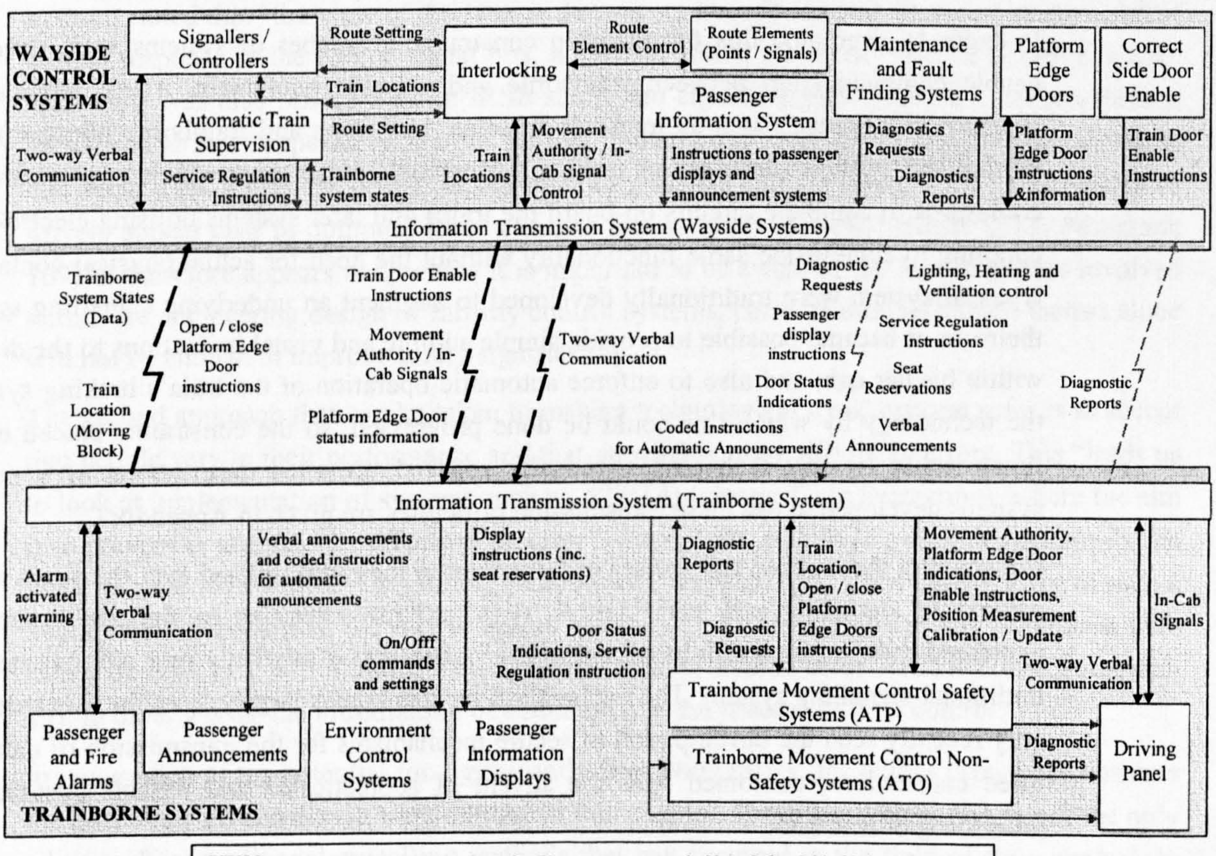
As with most signalling related terms, transmission based signalling does not mean exactly the same thing to all people. Generally, the term has been adopted to indicate the type of transmission and the way in which it is used. Indeed, it could be argued that the word 'transmission' is ambiguous and that its use has been misunderstood as a result. From this perspective a clearer understanding would be given by 'radio transmission' or 'radio communication'. This is a view supported by many definitions, such as:

- "A method of signalling where the information is provided to the trains by radio and where feedback from the trains is also received by radio transmission" (Schmid 1999, p7);
- The use of 'communication systems to give movement authority to the train without the need for lineside signals' (De Vilder 1998, p1).

However, the term 'transmission based signalling' is now widely used and a grasp of its meaning must therefore be developed. To this end, it can be noted that the common factor within all definitions is not the transmission technology itself, or even the exact way in which that technology is used. It is rather the opportunity that now exists for data transmission to be "leading and facilitating system development, not constraining it as in the past" (Lockyear et al. 1996, pC4/2).

4.2.1 *TRANSMISSION TECHNOLOGY*

It is possible to utilise several different mechanisms and transmission media in the communication of information between wayside control systems (with fixed locations) and moving trainborne control systems. Due to the complexities of modern transmission based signalling systems, and the different types of information that can be transmitted to the train (see Figure **4-1** for some examples), several technologies may be used, even within the same system. This section therefore outlines some of the main mechanisms and transmission media



used within transmission based signalling systems.

Figure 4-1: Wayside / Trainborne System Information Transmission

4.2.1.1 *LOCALISED TRANSMISSIONS*

Traditionally, devices that facilitate communication at a fixed point have been referred to as beacons, but in modern European terminology are more usually termed 'balises'. Such systems (see Appendix C3.2 for an example). Alternatively, the transmission mechanism may be contact free, by means of magnetic induction (see for example Appendix C3.7) or electromagnetic radiation (see Appendix C6.1). In each case, the train is able to detect its transit over the balise, but is only able to receive transmissions from the balise for the duration of that transit. This means that the balises can act as location identifiers, but that only relatively short messages can be reliably communicated by each balise. For example, the Eurobalise used for ETCS systems is designed to transmit telegrams of up to 1023 bits (Reddy 2001).

In order to extend the part of a route over which messages can be received, and the length of messages that can be transmitted, an inductive loop can be used. This consists of an insulated conductor laid as a loop between the rails in the area over which communication is required. Whilst a train is over the area covered by the loop, an antenna on the train can then receive signals transmitted in the loop.

In many respects, provision of an inductive loop enables the transmission system to closely replicate the performance that could be achieved by a traditional line side signal, by providing continuity in signal coverage on approach to a signal. "Although conventional signalling usually consists of widely spaced signal aspects, considerable continuity is achieved through the medium of the driver's eye" (Weber 1975, p2). A system relying purely on balise transmission, however, would have inherently less flexibility in update locations than the conventional signal / driver interface.

4.2.1.2 CODED TRACK CIRCUITS

The use of coded track circuits came about due to the need for continuous control systems to permit speed limit enforcement in accordance with the conditions imposed by multi-aspect signal layouts. When a track circuit is fed with an alternating current, this can be detected by coils on the front of the train (as long as they are located between the point at which the current is fed to the track and the first axle that will short circuit the track). By this arrangement, the presence of a train short circuits the coded signal for any following train, thus indicating danger as long as the absence of a signal is interpreted as the presence of a train. The coded signal is therefore provided in a fail-safe manner (Barwell 1983, p111).

At first, codes were applied by superimposing an a.c. current onto a D.C. track circuit. The D.C. current was then used to control wayside equipment and the ac signal to control in cab equipment. As system development progressed, it was recognised that the coding signal in a coded track circuit could itself operate wayside track relays via suitable code detection equipment. The first system to be used in accordance with this approach was installed at Philadelphia in 1933, by the Pennsylvania Railroad of America (Duckitt 1967, p1). This system utilised a current generated by applying an appropriate a.c. voltage at 3Hz for a clear (green) indication, 2Hz for a medium approach (double yellow) and 1.25Hz for approach (yellow), with no voltage being applied for stop (red) (Barwell 1983, pp112-3).

In general, coding can be applied to a track circuit as either frequency coding (where 'on' and 'off' code pulses are of equal length, with that length varying according to the code), or as time coding (where the cycle length is constant, and the code is given by the proportion of the cycle for which the code is 'on'). The carrier can also be either a.c. or D.C. (although where D.C. is used the ballast can sometimes act like a capacitor and store energy, so that the 'off' pulse voltage does not actually fall to zero) (Duckitt 1967, p2-4). It is also possible to configure coded track circuits to be reversible, where during the 'off' cycle, a return code is sent from the relay end to the feed end. This allows the track circuit to be used for movements commencing at either end without risking loss of the signal (Duckitt 1967, pp11-13).

Coded track circuits have been used on numerous systems throughout the world. Amongst recent examples is London Underground's Central Line (see Appendix C7.5), which utilises audio frequency jointless track circuits (for operation of the track circuit) superimposed with lower frequencies for transmission of control signals to the train (Jeffrey 1999, p2).

The continuous nature of the coded track circuit signal means that a change in signal aspect can immediately be passed on to trainborne equipment, overcoming the major limitation of localised transmission systems. However, the range over which coded track circuits can operate and the rate at which information can be transmitted over them are limited due to the attenuation characteristics of the rail, making them inadequate for use in many modern systems (Barwell 1983, p114; Fenner 2002, p32; Uebel 1991, p166-7).

Coded track circuits must also be arranged so as to ensure safe operation under worst case train performance. In reality, however, most trains will be able to brake harder than a worst case train, causing them to reach the command speed earlier than is necessary. These trains will then be required to maintain speed for some distance before the next speed command is received, causing their average speed to be reduced and an increase in energy consumption. The effect of differences in performance between train types could be minimised by multiple code systems that mean different things to each train type, but such systems would still not be able to deal with variations in performance within the same train type (Nishinaga 1994, p3).

The bandwidth available for transmission through the running rails is limited by the low electrical resistance between them. For example, depending on the ballast resistance, the power loss experienced at 100kHz is between 14dB and 30dB per kilometre (Alston et al. 1971, p26).

4.2.1.3 CONTINUOUS CONDUCTORS

In this type of system, insulated conductors are laid along the centre of the track. The simplest use of this approach is exemplified within the Docklands Light Railway (DLR) signalling system (see Appendix C7.7), where two conductors are laid so that they transpose at regular (25m) intervals, as shown in Figure 4-2 (Lockyear 1998, p53). A detector on the train then receives an approximately sinusoidal waveform that corresponds with the pattern of the conductors as they cross over and can be used to assist in determining the train's position. The regular position corrections achievable by this approach allow odometry system errors to be minimised.

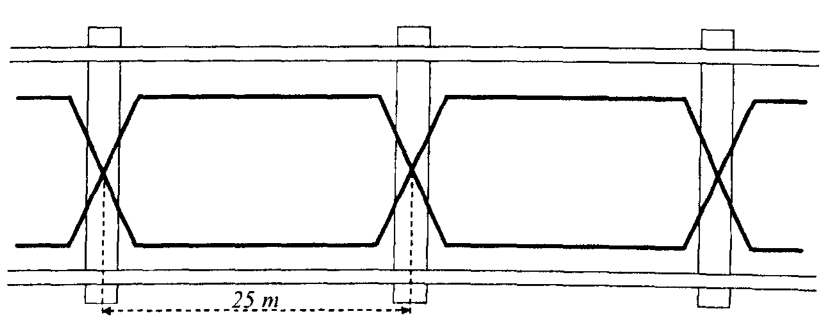


Figure 4-2: DLR Continuous Conductor Formations (Alcatel, 1996)

Systems using this technique have also been developed by both British Railways (BR) and Deutsche Bundesbahn (DB). The DB system, in use on all high speed lines in Germany, uses one conductor laid in the centre of the track and a second clipped to the foot of one rail. The two conductors are then transposed at regular intervals (as with the Docklands system already described). In the BR system, one of the conductors was laid straight and the other disposed in a triangular configuration, with the conductors still transposed at regular intervals (88 yds). As with the Docklands system, a high frequency carrier was then applied to the conductors (representing a safety signal), but it was also frequency modulated to provide variable information (such as signal aspects) to the train. In addition, the frequency of the triangular configuration in the conductors themselves could also be varied to provide a target speed code to announce speed restrictions to the train (Barwell 1983, pp114-117). An example of how this system might be configured is shown in Figure 4-3.

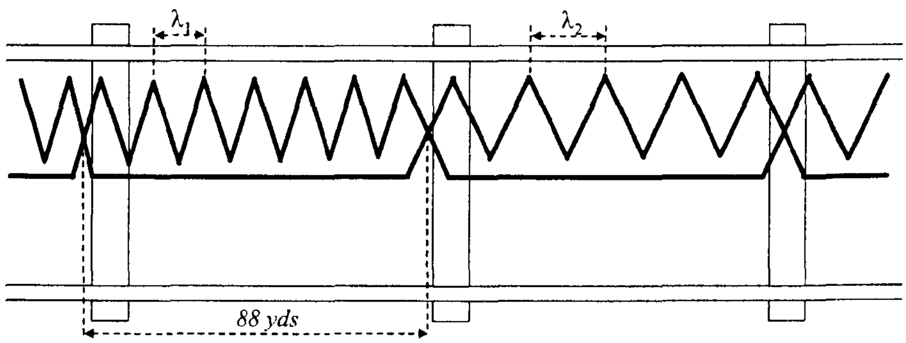


Figure 4-3: BR Continuous Conductor Formation (Barwell 1983, pp114 & 116)

Systems utilising continuous conductors can achieve high levels of data transfer, e.g., up to 1.2 Kbits per second for the DLR SETRAC system (Alcatel 1996, p7). Whilst this rate is not as high as that achievable by balises (up to 564.48 Kbits/s for an ETCS Eurobalise), the transfer is not limited to discrete locations, but can be achieved over the entire line (Reddy 2001).

An obvious drawback of continuous conductor approaches is the cost involved in installing the conductors throughout the railway. The cost of fixing the cable in the 'four foot' was found by BR to be a major factor in the overall cost of the complete track loop (Alston et al. 1971, p28). In addition to this, continuous conductors can cause maintenance difficulties, with the need to remove them prior to tamping the track bed and to replace them afterwards. In general, they are vulnerable to damage, wear and tear, vandalism or even theft (Tomlinson 2001, p34 & 36; Wright 2004, p12).

Attenuation in track conductor loops is better that that for coded track circuits, but still dependent on ballast resistance. As an example of this, early work by BR on track conductors showed that in 1000V double insulated cable the attenuation at 100 kHz varies between 5dB/km in dry conditions and 16dB/km in wet conditions (Alston et al. 1971, p27).

4.2.1.4 RADIO COMMUNICATION

The first experiments in providing radio communications between moving trains occurred in Britain in 1923 (Rumble et al. 1993, p1). Since then, continued developments in radio transmission technologies have led to radio becoming probably the most obvious form of transmission to trains. The term 'radio' is, however, a broad one that encompasses a number of transmission media. For railway applications, the range of choices would include:

- Low VHF (70-88MHz): Wide area coverage between ground staff and ground to train (shunting yards). Particularly used where there is no electric traction, due to the frequencies being susceptible to electric traction interference.

- High VHF (155-220MHz): The most effective for wide area coverage, being less affected by electric traction interference. It is also less affected by irregular propagation conditions and is hence used for trunk radio systems such as the National Radio Network (NRN) for British railways, operating in the 200Mhz band.
- UHF (420-470MHz): These frequencies are used for local area communication schemes including train to signal box communications. They are less susceptible to traction and electrical interference. The ORE recommends this as the best radio band for communications with moving trains due to the topographical layout of railways, i.e. cuttings, tunnels, embankments and open areas. The UIC Cab Secure standard specifies use of the 450MHz band, whilst the BR Cab secure system actually operates in the 460MHz band.
- Higher UHF (900MHz): Predicted for widespread use throughout Europe by the early 21st century. The GSM-R radio network that will carry voice and data transmissions for ETCS systems has been allocated this frequency band (with an up link of 876 to 880MHz and down link of 921 to 925MHz).

(O'Neil 1999, p134; Rumble et al. 1993, p1; Watkins(1) 1999, pp3-7; Watkins(2) 1999, p3).

The choice between radio frequencies is only partly down to the relative advantages and disadvantages of each for the role required. In most countries, radio frequencies are licensed and allocated by government agencies, in order to control interference (Rumble et al. 1993, p1) and to generate income. In order to address the problems associated with frequency allocation, a further transmission medium has been developed which also has potential uses within the railway environment:

- Spread spectrum radio: Where the data signal is spread over a bandwidth much wider than that of the original information, by use of a pseudo-random number sequence which is independent of the information itself. The pseudo-random number sequence has the same characteristics as white noise, but can be exactly reproduced (Rodgers 1995, p4).

Spread spectrum transmission has a number of advantages. In particular, the transmitted signal does not require allocated bandwidth and is immune to electro-magnetic interference. It also contains inherent time referencing within the spreading signal. Time demarcations can be established by synchronising to this signal and then used to compare the time of transmission with that of arrival. Given that the elapsed time equates to the distance from the transmitter (at sea level the speed of radio propagation is a constant 0.2998m per nanosecond), a single transmission is sufficient to calculate a train's location. The accuracy of this calculation depends on the spreading and sampling rates (improving as they increase), as well as the number of transmissions that are analysed. For example, combining a 5MHz spreading rate and 16 times over-sampling in the receiver correlators provides a range resolution of 12 feet on any one individual transmission. Pairing the measurements for an outbound and return transmission between any two units and averaging the results eliminates the effects of any offsets or drift, and increases the resolution to 6 feet. If each transmitting unit communicates to at least 2 other units, the measurements recorded by and from each can be compared, giving a location accuracy of 3 feet. Thus, any desired resolution can be achieved by increasing the number of transmitters along the track. This in turn means that the trains do not require such sophisticated odometry systems and that most equipment can be cab mounted and self contained, rather than mounted on the train's bogies (Nishinaga et al. 1994, p15; Rodgers 1995, p5).

There are also a number of possible mechanisms for transmission of radio signals. The simplest of these would be free space transmission, as typically used for commercial television and radio broadcasts, where the transmission is sent in an unfocussed manner over a wide area. This approach is commonly used for general voice radio communication, and in modern 'train-order' system approaches to signalling (see section 3.6). However, it relies upon good reception wherever needed. Where the transmission of data is to occur, and particularly where the train systems rely upon correct and timely receipt of that data, more sophisticated systems of transmission are usually required.

There are a number of mechanisms that can be utilised to provide a more focused transmission of the radio signal, including:

- Designing transmitters to focus along the line of the track. Under this approach, the transmission is still in 'free space' and able to cover a large area including multiple trains. In tunnels the term 'distributed antennas' is commonly used for this approach. The design of such systems must ensure that sufficient antennas are provided for the radio signal to propagate to all areas of the tunnel that require coverage. It must also account for the fact that the transmission will only be successful if there is not an obstruction (such as a second train) in between the antenna and the intended recipient train;
- The use of leaky feeder cables. These use perforated coaxial cable laid alongside the track, designed to allow the signal to radiate into the air (and thus set up a radio signal). Tests on the Washington D.C. Metro have shown that, at 800 MHz, both distributed antenna and leaky feeder systems offer acceptable coverage in a railway tunnel environment, with the leaky feeder system offering cost advantages (Jakubowski 1994, pp1113-4; Richardson 1999, p46);
- Further improvements for focused radio transmission have been claimed by the 'IAGO' waveguide, as used for microwave transmission of digital Direct Sequence Spread Spectrum messages on the Singapore North East Line (Alstom, 1999; Chew et al 2001, pp2-3).

Radio transmissions can be used within transmission based signalling systems to provide coverage of the whole, or parts of the railway line. Where the whole line is covered the system can make use of continuous transmission, with the possibility of transmission from track to train or train to track at any time. However, if the radio coverage is only provided in certain places (such as in or on the approach to stations, or on the approach to conventional signal locations) then it is only possible to support intermittent transmission (albeit at higher rates and longer distances than that supported by balises or loops).

4.2.2 IN CAB SIGNALLING

Linked with the concept of transmission based signalling is the idea of cab signalling. As the speed of operation is increased on a railway, it becomes difficult for the driver to observe and react in a safe manner to conventional lineside signals. In order to overcome these difficulties and support operation above 200km/h, transmission systems can be used to provide information in the cab that the driver can then follow in place of, or in conjunction with, any signal aspects outside of the cab. In-cab signalling acts as an enabler for the more advanced method of train separation known as moving block control (see section 4.2.4). It is also often an integral part of train control systems, such as Automatic Train Protection (ATP) systems.

The term 'in-cab signalling' is usually used to refer to systems that provide driving information within the cab, rather than via fixed signals along the lineside. In accordance with this, the definition of in-cab signalling given by the International Electrotechnical Vocabulary is:

> A "fail safe system giving indications in the driver's cab of the situation of the line ahead such as speed limits, signal aspects, distances to speed restriction points, and target speeds" (BSI 1998, p124).

Where in-cab signalling is provided as a part of an ATP system, the driver becomes influenced by the indications that it provides. This means that the ATP system becomes in effect the only safety system for the train, rather than an additional safety system (as it would be if the driver remained independent of the ATP by basing his/her actions on a separate signalling system - whether in-cab or trackside). Under such circumstances, the design criteria for the ATP system must be different to those that would be required for an independent safety backup to the driver. It should, therefore, be remembered that whilst ATP fitment could well act as an enabler for cab signalling (in that cab signalling may be justified by the approach being taken to ATP), it is not a prerequisite of ATP. Also, where both systems are provided, they do not necessarily have to be integrated.

4.2.3 FIXED BLOCK CONTROL

The subject of conventional fixed block signalling, with 2, 3, 4 (or more) aspect signals located along the lineside has already been discussed in section 3.11. With transmission-based techniques, high-resolution data with fast update rates can be passed from the trackside to the trainborne systems. This enables complex brake assurance systems and in-cab signalling to operate on the train (Riley 1999, p2). It, therefore, becomes possible to consider removing the conventional lineside signals and to operate an 'n' aspect system with an in-cab display.

In section 3.11.5, it was noted that fixed lineside signals displaying 3 or more aspects are required to be separated by approximately equal distances in order to avoid driver confusion. It was also noted that throughput is increased as the signal separation is reduced, but that safety demands a minimum separation of at least the trains braking distance from line speed between 'n-1' consecutive signals. This means that for maximum throughput of trains, the signal separation must be reduced when more aspects are used. The use of additional aspects with lineside signals therefore requires additional equipment at each signal location (for the extra aspect) and an increase in the number of signal locations as the block lengths separating them is reduced. Implementing this approach with in-cab displays therefore represents a more cost effective solution than lineside signalling, requiring far less trackside equipment.

In practice, however, replicating the use of an 'n' aspect approach by reducing block lengths still requires the use of additional trackside equipment for each 'aspect' added, in order to provide delineation between blocks. If it is further considered that the headway benefit obtained by the use of an extra aspect reduces with each increase in aspect numbers (see section 3.11.5), it can be seen that increasing the effective number of aspects in this way is still not an ideal solution to the problem of improving headway. Fortunately, it is also not the only way of increasing throughput by use of in-cab signals. In section 3.11.2 the use of 2-aspect fixed block signalling (which does not require blocks of equal lengths) was considered. At speeds requiring the use of a repeater for sufficient sighting of the signal at line speed, two main limitations were identified for the application of this approach with lineside signalling:

1. As the signal spacing approaches the trains braking distance, the amount of trackside equipment required is more than that for a 3 aspect scheme with equivalent signal spacing;
2. If the separation of signals is closer together than the trains braking distance, repeater signals must be located before the preceding signal. This gives rise to the possibility of driver confusion in interpreting which signal is being repeated;

With the use of an in-cab signalling system these limitations can be overcome by designing the system to interpret the conditions ahead into a simple in-cab display to the driver. This means that the lengths of individual blocks do not need to be kept equal, but can be optimised in the same way that 2-aspect signal separations can in conventional fixed block signalling. Adopting this approach can further minimise the system complexity and costs.

4.2.4 MOVING BLOCK CONTROL

With all fixed block systems, the train position resolution is limited by the length of the block sections. As a consequence of this, the capacity of the railway is constrained by the length of the signalling blocks (Lockyear et. al., 1996). However, shortening track circuits to allow for increased capacity can become:

a) prohibitively expensive;
b) a serious reliability problem due to increased amounts of equipment;
c) operationally crippling on existing systems due to the amount of trackside installation. (Nishinaga 1994, p3).

In addition to this, fixed block signalling must be designed for an assumed train performance and service pattern, and the optimal capacity can therefore only be achieved when those assumptions are met in practice. This means that a train can only be operated at higher speeds than the design speed if the train path ahead of it is kept clear (reducing overall capacity by a train path). Similarly, if a train can only operate at lower than design speeds, it will occupy more than one designed path whilst travelling through the section (reducing capacity and potentially delaying trains behind it). If a train's braking is worse than that assumed in the design, it can only be permitted to travel through the section at lower speeds, whilst if its braking performance is better than that assumed in design it will be constrained to maintain a greater separation from the proceeding train than could safely be allowed.

The limits of fixed block approaches have long been recognised and as early as 1938, a proposal for overcoming them was made to the IRSE:

> "as a possible method of decreasing headway some form of cab signalling, which is dependent on the relative speeds of following trains, might be developed and be such that trains could always run with just sufficient braking distance between them. In other words, the block sections would move with the trains" (Woodbridge 1938, p198).

This idea is now commonly referred to as moving block control. A number of systems now exist, or are currently in development, world wide, utilising transmission technology to achieve a moving block. These systems generally have onboard equipment that enables trains to determine their position by sensing widely spaced fixed markers along the track and counting wheel revolutions. Alternatively, use is also made of GPS, doppler radar, continuous conductor transpositions and spread spectrum radio to determine position. Once each train's position is known, continuous (or pseudo-continuous) communication with a central control system (or in theory, with other trains) allows safe movement limits to be determined.

As this approach does not require the use of track circuits (or equivalent track based infrastructure) in determining trains' positions, it further reduces the need for track based equipment and therefore offers the opportunity to reduce infrastructure costs and maintenance access problems. The removal of track based train detection does, however, raise new problems. These are primarily related to:

- Train integrity (each train now needing to prove that it is complete in addition to determining where it is) and;
- Rail integrity (with the loss of the limited broken rail detection offered by track circuits).

In section 3.11.6 it was shown that an 'n' aspect signalling system would have a headway time given by :

$$Ht_n = \frac{(n-1)d + S + O + L}{V} \qquad \text{(where n>2)} \qquad \text{Equation 3-11}$$

It was also noted in section 3.11.5 that for a calculated braking distance (B), the best possible headway for an 'n' aspect fixed block signalling arrangement (where n>2) would be when the average signal spacing (d) was equal to B/(n-2). This means that:

$$Ht_n = \frac{\left[\dfrac{(n-1)B}{n-2}\right] + S + O + L}{V} \qquad \text{(where n>2)} \qquad \text{Equation 4-1}$$

As 'n' tends towards infinity, the signal separation becomes smaller, and hence the positional accuracy becomes greater. Ultimately, the block lengths become infinitely small, and the system has in effect become a moving block system. This means that the theoretical headway time under moving block operation is in fact the same as the headway time for a fixed block signalling system with an infinite number of aspects:

$$Ht_\infty \Rightarrow \frac{B + S + O + L}{V} \qquad \text{Equation 4-2}$$

This equation can also be derived from first principles for a moving block system, as shown in Figure 4-4.

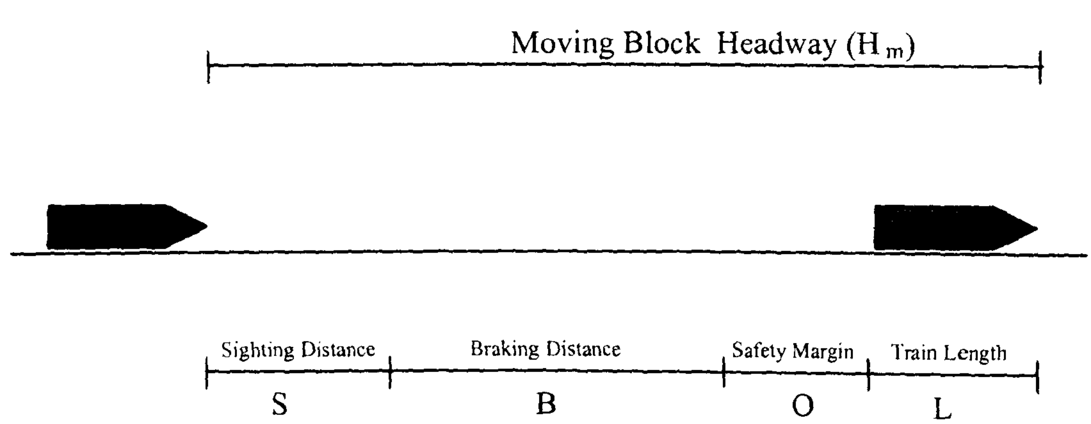


Figure 4-4: Moving Block Headway

Here it should be noted that as the position of each train must be reported by the train itself, there will always be an element of uncertainty as to the trains real time position based on the system's position determination errors. There will also be a delay within the position reporting system itself. Although not mentioned in section 3.11, delays and accuracy errors such as these occur in all systems, and will effect the achievable capacity. In practice, they must be allowed for when determining the sighting distance that applies to the system (e.g it is not just driver reaction but also system reaction that contributes to the actual sighting distance). In the case of moving block systems, which must use in-cab signalling displays, it is also likely that the time required by a driver to observe a change in display will be different to that of fixed line-side signalling displays. This must also be allowed for in determining sighting distances.

It has been claimed that moving-block systems can increase capacity by 30 to 50% on that achievable with a conventional fixed block-signalling system (Wang et.al. 1993, p153). However, there is controversy about this. Some authors note that railway systems always contain elements of a fixed block nature (such as junctions, stations, ventilation points and speed restrictions), which will all tend to reduce the benefits that the moving block system offers (Clark 1999, p7). Others claim that the major advantage of moving block operation arises at junctions or in stations, when trains have to follow each other at short intervals (Schmid [1] et al. 2002, p6). What is clear, is that moving block systems have the ability to optimise train throughput to the actual train performance and separation encountered. This ability should offer improvements in system recovery from states of congestion and irregular working when train speeds become abnormally low (Barwell 1983, p47; Schmid [1] et al. 2002, p6).

There also appears to be disagreement between experts as to whether the greatest benefits of moving block systems are obtained on railways where all stock have the same performance and stopping patterns, or on railways operating mixed stock and stopping patterns. Whilst the author has witnessed and been involved in many animated discussions of this issue, he has been unable to find any published material relating to this debate.

4.3 AUTOMATION IN TRAIN CONTROL SYSTEMS

The terminology used to define the elements of automation that are incorporated within Train Control varies from country to country, company to company and author to author. However, it is the intention of the author of this report to adopt the use of terminology that has become most prevalent in the UK in recent years.

Whilst a number of terms related to the automation of train control are widely used within published material, references to the definition of those terms are fairly rare. Where definitions are given, most authors select a sub-set, often focusing upon the definition and explanation of Automatic Train Control (ATC), Automatic Train Protection (ATP) and Automatic Train Operation (ATO) systems. Where a broader set of terms are used, it is common to refer to either Automatic Train Supervision (ATS) or Automatic Train Regulation (ATR), but usually not both. Automatic Route Setting (ARS) and Automatic Traffic Monitoring may also be referred to (see Waller 1975, p229; Doherty 1995, p9; Barton et al. 1998; Basu 1998, p78; p1 Dapre 1999, p16 and Goodman[1] 1999, p1 for a selection of definitions). However, the author of this report has not found any published material that uses, let alone defines, all of these terms and their interactions or hierarchy in one place. In order to piece together the fragments of information that are available from published sources, a series of interviews were conducted with experienced professionals within the fields of Railway Signalling and Operation (see Appendix A). From the responses obtained, the author was able to identify a general (although not unanimous) opinion that the terminology can best be represented as shown in Figure 4-5.

Most variations from this view can be traced to geographical differences in the usage of terms, or to changes in the meaning applied to those terms over time.

In the following sections the author will define the terminology used to describe railway automation systems used in Britain, in accordance with Figure 4-5. The author will also explain the alternative definitions that exist and, where possible, will identify the source of the differences.

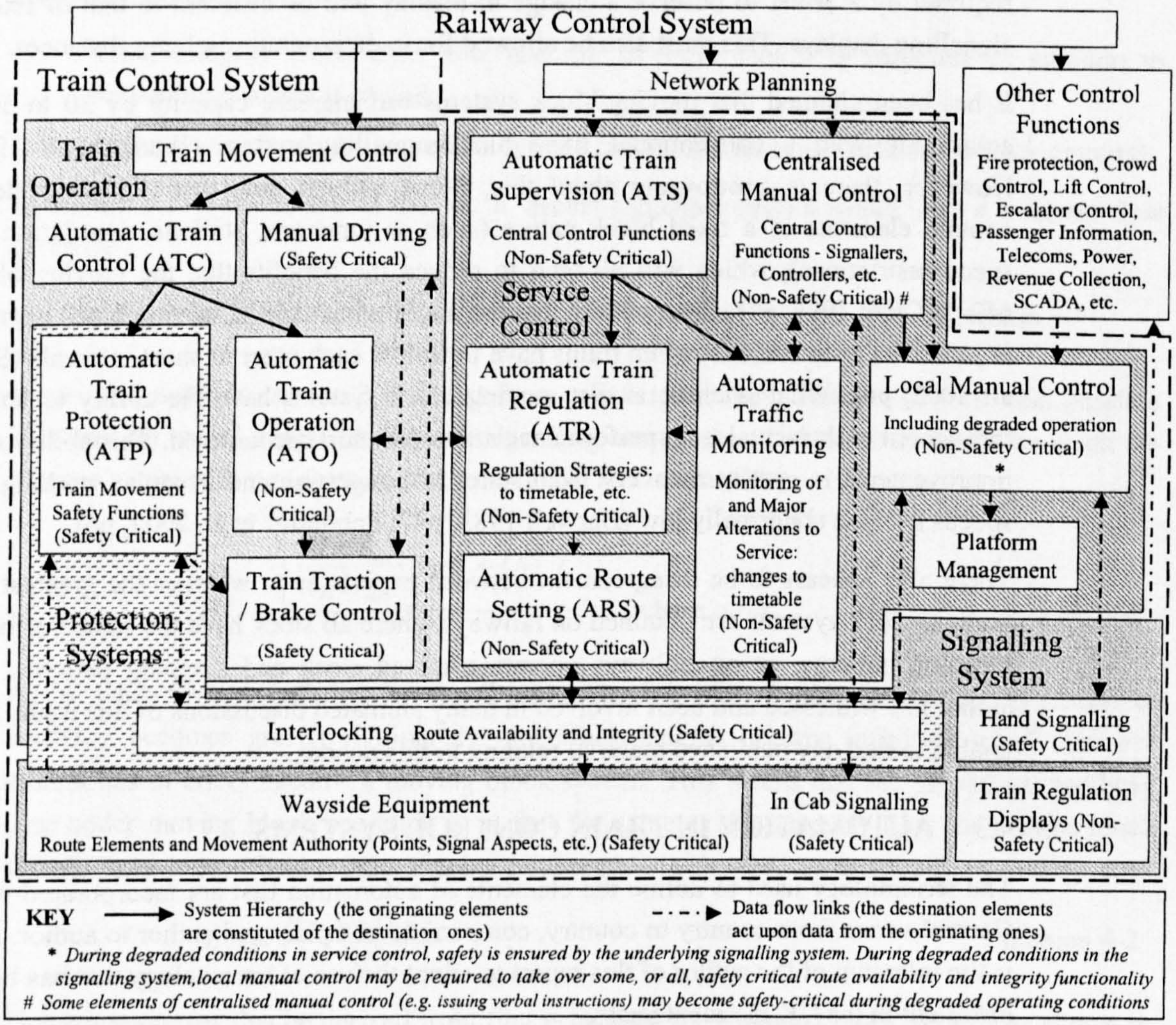


Figure 4-5: Automatic Train Control Systems Hierarchy

4.3.1 AUTOMATIC TRAIN CONTROL (ATC)

There are four main groupings of definition given to the term Automatic Train Control (ATC):

- Before nationalisation, some of the UK railways used to refer to in-cab warning systems as ATC, where they would now be more commonly considered as a basic form of Automatic Train Protection (ATP). This use of the term has been perpetuated in some places – for instance in Australia, where it now has the same meaning as ATP in the UK;
- A second view is that the term ATC applies to a system directly controlling a train's brake and traction interface, as a sub system of ATO. This appears to be an historical usage from the early days of automation and is not now in use by any of the system manufacturers. Adherence to this definition now appears to be in decline and is mainly adopted by mainland European engineers (see Appendix A, interviews with Eberhardt and Irwin).

- The third view is that ATC refers to all automation systems and is constituted of any combination of ATP, ATO, ATS, ATR and ARS. An example application of this view can be found on the Washington Metropolitan Area Transit Authority Metrorail system, which includes an ATC system consisting of ATP, ATO and ATS sub-systems (NTSB 1996, pp16-17). It was also expressed in some reference sources and in several of the interviews conducted (see Appendix A, interviews with Jerry Lewis and Marcus Eberhardt; Allen A 1996, p3; Gill 1986, p3/1 as examples), but was not found to be the majority view.
- The fourth view identified considered ATC as a collective term for systems automating the tasks of the train operator (or driver). That is, being constituted of ATP and / or ATO (Kitchenside 1998, p191; Pachl 2000, p3/25). This was found to represent the majority view in the UK.

Based upon the findings of his research the author proposes to adopt the fourth view, as represented in Figure 4-5, with the definition:

> **Automatic Train Control (ATC)** - 'a generic term for describing systems which reduce the reliance placed on a human driver for the safe and efficient movement of trains'.

Where ATC includes both ATP and ATO, it can utilise stored data and trackside / trainborne sensors to create real time instructions for the safe automatic control of an individual train's movements. It is not, however, necessary for an ATC system to include both sub systems. ATP may typically be in existence on its own.

Light rail and metro systems, where all trains have similar physical and performance characteristics, are ideal for the application of ATC. However, the typical mixed traffic scenario applicable to main line railways is more difficult to automate fully. As a result the application of ATC to main line railways has tended to be limited to the use of ATP functions, and then mainly on high speed lines (Goodman[(1)] 1999, pp1-2). The development of other ATC functions is now beginning to be considered for main line railways, as evidenced by the KOMPAS development projects and proposed Intermobil Region Dresden operational tests of ATO in Germany (Eberhardt 2001, p7). However, such projects are the exception rather than the rule.

ATC systems are generally credited with offering improvements in performance (through ATO) and safety (through ATP). This can in turn lead to cost benefits through greater utilisation of infrastructure assets and the potential to reduce staffing levels.

4.3.2 AUTOMATIC TRAIN PROTECTION (ATP)

It has already been noted in section 4.1 that the elimination of human error would require the removal of the humans from the system. In accordance with this, the IRSE recommend that by far the most effective means of minimising the consequences of human error is the provision of a system of automatic train protection (IRSE[(1)]2001 Annex D Paragraph 2). However, the definition of ATP varies considerably between authors, making the exact meaning of this recommendation ambiguous. As an example of this, two international definitions of ATP are:

> a "system using information of signal aspects, track speed limits, train speed supervision and driver reactions to automatically prevent a train passing a signal at danger and / or exceeding speed restrictions" (BSI, 1998);

and

"a safety system that enforces either compliance with or observation of restrictions and signal aspects by trains" (ERTMS Users Group[1] 1998).

This difference in definition leads to numerous systems being referred to as types of ATP, some of which are described in Appendix C of this report. In practice, most people now use the term ATP to refer to a comprehensive system that enforces both speed and movement authority and, on closer examination of the IRSE Review of Signalling Philosophy, it is this interpretation that is implied by the recommendation given at the start of this section (see IRSE[1] 2001). It is also the interpretation contained within railway group standards:

"The automatic mitigation of the consequences of failure of the driver of a train to comply with signal displays and permissible speed orders by means of a model of the permissible speed profile of the train, as determined by its characteristics and the signalling conditions and continuous sampling measurement of the train's speed and distance run" (Blakeney 2002, p5).

This is also the interpretation that is required by the UK Railway Safety Regulations, which define ATP as:

"Equipment which -

a) Causes the brakes of the train to apply automatically if the train-
   i) Passes without authority a stop signal such passing of which could cause the train to collide with another train, or
   ii) Travels at excessive speed on a relevant approach.
b) Automatically controls the speed of the train to ensure, as far as possible, that a stop signal is not passed without authority and that the permitted speed is not exceeded at any time throughout its journey" (Uff et al. 2001, p50).

In the opinion of the author, assuming that the term 'signal aspect' can be interpreted to mean 'movement authority' (and thus include systems that do not contain physical signals), this definition of ATP should become the one commonly used. However, as highlighted by the example definitions already given, this has not been the historically accepted definition and is not yet the only definition in use.

A further example of ambiguity in the definition of ATP can be seen in the Railway Safety Regulations themselves. Regulation 3 prohibits the operation of a railway in the UK over 25mph without ATP after 1 January 2004, a statement that on its own could be taken as a requirement for all train movements to be comprehensively protected by ATP by that date. However, under regulation 2 it is made clear that this is not the case. The ATP train stop function is only required to apply to signals protecting junctions and not those on plain track, whilst the ATP speed trap functions are required only for:

- An approach to a stop signals (unless the trainstop system would stop a train at maximum speed before the fouling point without the need for a speed trap)
- An approach to speed restrictions (beyond specific limits) and
- An approach to buffer stops

(Uff et al. 2001, p51-52).

This demonstrates that not only the type of protection provided varies between definitions, but also the extent of application.

Considering the scope available to interpretation when the term 'ATP' is used, there is a clear need to clarify the definition that will be assumed within the remainder of this report. However, the author would contend that, due to the historical variation in use of the term, it is not now adequate to simply define ATP. Instead, a series of terms and their definitions are required in order to differentiate between the possible interpretations. To this end, the author proposes the following series of definitions:

Train Protection - 'the generic name for all systems designed to assist **or** enforce compliance with **or** observation of **some** or all speed restrictions **or** movement authorities. Train protection systems may be applied **either** manually **or** automatically';

Warning Systems - 'systems assisting observation of movement authorities, based upon manual activation';

Automatic Train Protection - 'a system automatically enforcing compliance with **or** observation of **some** or all speed restrictions **or** movement authorities'. This is the generic name for a train protection system that applies automatically;

Automatic Warning Systems - 'systems automatically assisting observation of movement authorities';

Automatic Train Stop - 'a system automatically enforcing compliance with movement authorities';

Automatic Speed Supervision - 'a system automatically enforcing compliance with speed retrictions'. An example being TASS

Partial ATP - 'a system automatically enforcing compliance with speed restrictions **and** movement authorities at **some** locations or for **some** vehicles';

Comprehensive ATP - 'a system automatically enforcing compliance with **all** speed restrictions **and** movement authorities (for all vehicles) within a given area'. This type of system is often divided into two sub-categories, Intermittent ATP and Continuous ATP. The differences between these two sub-categories are detailed in Appendix C.

4.3.2.1 ATP FUNCTIONALITY

In order to provide full supervision of trains with respect to speed restrictions and signal aspects, a comprehensive ATP system requires information relating to:

- **Dynamic data:** the current train speed and location, brake pipe pressure (for service & emergency braking), and master controller position;
- **Train data:** the class, length, acceleration performance, braking performance (for service and emergency braking) and maximum permitted speed of the train;
- **Route data:** gradients, current maximum line speed, the line speed profile ahead (relevant to the particular class of train) including the start and finish points of temporary speed restrictions, the distance to the next signal/marker/data transmission point, the distance to go before the train must slow down or stop (the movement authority).

  (Barnard 1991, p294; Dapre 1999, p9; Rose et al. 1989, p2)

In addition, knowledge of track adhesion conditions can assist the system in optimising the approach to braking. The system may also require inputs from the driver to allow adjustment of the parameters on which it operates and to acknowledge warnings.

In response to this information, the ATP system must have a fail-safe output to control the train's brakes. It may also have outputs in the form of advisory alarms and indications to the driver (typically train speed, the maximum current speed limit, the target speed at the next signal/marker, warning of impending brake intervention and an indication of when brake intervention has already occurred). However, if the ATP system is provided for safety only, there is no need for visual assistance prior to corrective action (Rose et al. 1989, p2).

Numerous ancillary outputs could potentially be added to this list, including operation of the train horn, operation of lights, control of heaters, inputs to 'black box' recorder units, raising and lowering of pantograph and other train management functions (Rose et al.1989, p3). It is also common for ATP systems to control features such as correct side door enable. However, where such ancillary outputs are required it may be more appropriate to consider them as a part of an associated ATO system.

Some of the required information will be fixed for a particular train or route, and can therefore be stored in either the trainborne memory of the ATP system or in a trackside memory store for transmission to the train as required. Other information, such as speed, acceleration and train locations, must be determined by means of sensors on the train.

Location information is generally relative to the last passed track circuit boundary or transmitting beacon location, but may be relative to a single fixed point for the entire railway.

Comprehensive ATP systems continually calculate the maximum safe speed for the train to travel at and monitor actual train speed against this. If the actual speed begins to exceed the maximum safe speed most ATP system will give a warning to the driver and, if no action is taken to bring the speed back to acceptable levels, automatically apply the train's service brakes. Where this is inadequate to reduce speed sufficiently or stop the train quickly enough, the emergency brake will then be applied. Due to the desire to avoid passenger discomfort and additional wear and tear, the emergency brake is not usually used as the first form of intervention – although an ATP system could operate solely on the basis of emergency brake control.

The inclusion of comprehensive ATP enables the safety systems to operate as a closed-loop in order to bypass the human driver if he makes an error. If ATP is provided without ATO, observation of signal aspects and the feedback loop determining the error that exists between that target and the train's actual movement are left to the driver during normal operation. This is represented in Figure 4-6.

It should be noted that, depending on the type of ATP system used, the driver might be provided with an indication of the ATP targets in the form of target speeds, distance to go, signal aspects, or any other form deemed appropriate. Indication of the train's current position and details of the line profile ahead may also be presented to the driver by the ATP system. For the sake of simplicity, however, these situations have not been included in Figure 4-6. Where there is no ATP system, the driver of the train is the primary safety system for train movements. When an ATP system is introduced, if the driver is allowed to continue controlling the train in the same manner as before, he can continue to act as the primary safety system for the train's movements whilst the ATP acts as a secondary safety system. This is the scenario represented in Figure 4-6 and also that traditionally adopted by London Transport in their automation of train control functions (Maxwell 1975, p230). If, however, indications are provided to the driver by the ATP system (as, for example, will be the case in the ERTMS systems), those indications influence the way in which the driver controls the train. The ATP system then becomes the primary safety system for train movements (see Appendix A, Interview with Jim Carpenter; Cook 2002 p3).

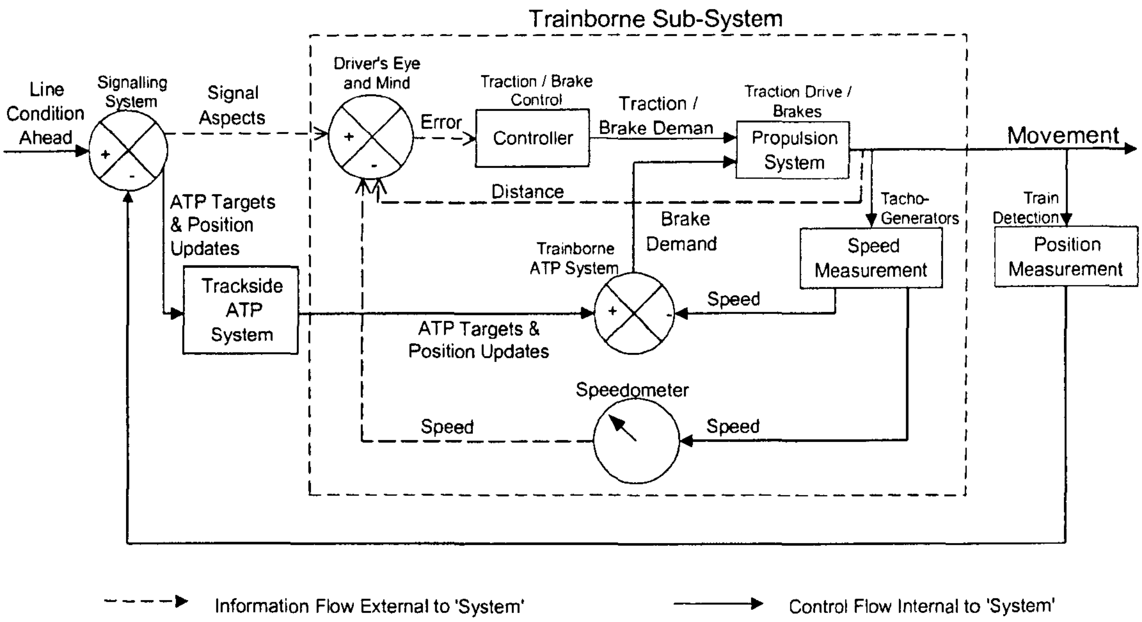


Figure 4-6: Train Control System with Automatically Closed Safety Loop (see Appendix B)

Not all of the functionality shown in Figure 4-6 would be provided by ATP systems that were not comprehensive (for example a trainstop system's only interaction with the trainborne sub-system is a brake demand). Along with these variations in the scope of protection afforded by the ATP systems, the fact that the data required for their operation can be obtained and formatted in a number of different ways is reflected in the range of different systems used by railways world-wide. However, despite the variety, the functional integrity of all ATP systems is vital and failure to provide that integrity would directly affect the safety of the railway (Jeffries 1991, p201). A more detailed discussion of the functionality of different types of ATP system is provided in Appendix C.

Whilst ATP systems offer improvements in safety by eliminating many forms of driver error that may result in SPADs and exceeding permissible speeds, it is important to note that they are not a cure for all of the causes of overruns and over speeding. For instance, problems of low rail adhesion can actually become worse under ATP if a full emergency brake application is invoked (IRSE[3] 2001 section 9.6). As a result of this, "the provision of ATP must not be used as an excuse for not making improvements in signalling systems that should reduce the likleyhood of driver error in the first place" (Cooksey 2001, p4).

It is also worth noting that crude ATP systems can come at the cost of reduced capacity, although more advanced systems can see capacity improvements (Goodman[1] 1999, p2).

4.3.3 AUTOMATIC TRAIN OPERATION (ATO)

During the 1960s, the term Automatic Train Operation was used in the same way that ATC is now used, to cover the whole field of automating the operation of a train. This is a definition that is still used by some people today, but once again the author of this report prefers to adopt the use of terminology that has become most prevalent in the UK in recent years. Under this use, ATO refers to the automatic control of train movements without the need for intervention of a driver (Taskin et al 1995, pA3/14). Hence the definition can be given as:

Automatic Train Operation (ATO) – "a method of operation in which the movement of the train is automatically controlled without the intervention of a driver, who, if provided, exercises only a supervisory function." (BSI, 1998).

4.3.3.1 ATO FUNCTIONALITY

It is important to note that the function of driving a train is safety critical. However, ATO is not a safety system but, as shown in Figure 4-5, embodies only the movement control aspects of the driving function. ATO cannot therefore exist without ATP, since it relies upon ATP to provide the movement safety functions (Taskin et al. 1995, pA3/14; Waller 1975 p243). As a result, ATO systems are not generally designed to be fail-safe, since the fail-safe ATP will override the ATO system in case of failure – just as it would a human driver. Where automatic train operation is to occur, there may as a result be additional safety functions that must be added to the supervising ATP system. These may include tasks such as obstacle detection and door control.

In order to enable automatic driving, ATO systems require the ability to perform a number of functions, including initialisation of operation together with the subsequent acceleration, braking and (where required) coasting of the train in order to maintain target speeds and manage scheduled stops (Taskin et al 1995, pA3/14; Eberhardt 2001, p5). They may also perform ancillary functions such as operation of the train horn, operation of lights, control of heaters, inputs to 'black box' recorder units, raising and lowering of pantograph and other train management functions (as discussed in section 4.3.2).

This means that they require information relating to:

- **Dynamic data:** the current train speed / location and brake pipe pressure (for service braking);
- **Train data:** the class, length, acceleration performance, braking performance (for service braking) and maximum permitted speed of the train;
- **Route data:** gradients, current maximum line speed, the line speed profile ahead (relevant to the particular class of train) including the start and finish points of temporary speed restrictions, the distance to the next signal/marker/data transmission point, the authorised direction of travel, whether or not coasting is permitted (or alternatively enough timetabling information to determine this for itself), the distance to go before the train must slow down or stop (the movement authority and the scheduled stopping pattern);
- **Control inputs:** any additional data required to control train management functions (these may, or may not, be specific to locations on the route)

As with an ATP system, additional knowledge of adhesion conditions can assist the system in optimising the approach to braking. All of this information must be acted upon in a manner that avoids intervention by the ATP system. The ATO system must therefore have an output to control the train's brakes. This need not be fail safe (as safety is ensured by the ATP system), but should also not be designed to operate in a manner that continually relies upon the ATP to provide brake intervention.

Some of the required information will again be fixed for a particular train or route, and therefore can be stored in either the trainborne memory of the ATO system or in a trackside memory store for transmission to the train as required. Other information, such as speed, acceleration and train locations, must be determined by means of sensors on the train. This information may be shared with the ATP system in the same way as a manual driver can be provided with in-cab signalling information by the ATP system. As in the case of manual driving, however, additional levels of safety can be achieved by designing the ATP and ATO as independent systems with independent data sources, so as to eliminate some potential common mode failures.

4.3.3.2 DRIVERLESS ATO

The basic ATO functions already listed would be sufficient for a driver accompanied system where, in the event of any abnormal situation, a human driver can intervene to prevent a collision or other incident from occurring. However, ATO systems can in fact exist in three distinct forms: driver accompanied, train captain operated and unaccompanied. (Eberhardt 2001,p2).

In the case of driver accompanied ATO, the driver remains in the cab and is therefore able to monitor the system's operation and the state of the track ahead. He/she is also able to perform station duties (door opening and closing) and to control the departure of the train from stations. Since the driver is located within the cab, he/she is in a position to intervene and override the system in case of emergency (such as observing an obstruction on the track) and to deal with any system failures. If necessary, the driver is available to drive the train manually or co-ordinate an evacuation. Where the train is driver accompanied, safety systems provided on manual trains, such as continuous radio communication with the central control and CCTV monitoring of stations with curved platforms, are generally still considered necessary (Lawrence 1982, p1/4).

Where the ATO system is train captain operated, it is usual for the train captain to be given other duties, such as checking tickets. This means that he/she is not available to provide continuous monitoring of the system's operation or the state of the track ahead. The train captain is still able to perform station duties (door opening and closing) and to control the departure of the train from stations. Since the train captain is still located within the train, he/she is also available to deal with any system failures. If necessary, this can include driving the train manually or co-ordinating an evacuation, but it cannot be guaranteed that he/she will be able to intervene and override the system in case of emergency (such as observing an obstruction on the track). This means that the system requires additional features, including:

- Surveillance of train operation and reporting of faults;
- Track observation, such as cameras to detect track discontinuities and platform cameras to detect passengers falling off (some parts of this functionality will be safety critical and fall to the ATP system).
- Protection mechanisms against guideway intrusion hazards (such as fencing, platform edge doors and CCTV monitoring).

  (Eberhardt 2001, p5; Jeffries 1991, p203-4 & 210; Lawrence 1982, p1/5).

In the case of unaccompanied ATO operation, there is no operator available to provide continuous monitoring of the system's operation or the state of the track ahead. The system must be able to cope with automatic door opening / closure and must control its own departures from stations (including reversing at terminals), unless platform based staff are available to perform these functions. Since there is no member of staff on board the train, the system must be able to deal with train or passenger movements under equipment failure or emergency conditions, and must be able to take action automatically in the case of emergencies (such as an obstruction on the track). In addition to all of the functions required of an ATO system under train captain operation, when unaccompanied the system also requires:

- To be able to communicate status reports to the control centre and raise alarms in the control centre if a train stops unexpectedly or problems occur;
- To receive back commands such as saloon lights control, passenger announcements, automatic coupling / uncoupling (if required), alarms and timetable revisions;
- Facilities for remote interrogation, to allow isolation of faulty equipment and slow speed operation under safe conditions when something goes wrong with trainborne equipment;
- To provide facilities for passenger announcements (either automatically or by control centre staff);
- To know the scheduled time of departure from a station;
- To provide 'passenger emergency button' type facilities to prevent a train from starting from a platform if a passenger is in difficulty (this functionality will be safety critical and may therefore be included as a part of the ATP system)
- Control of train doors (some parts of this functionality will be safety critical and fall to the ATP system)

  (Jeffries 1991, p202-4; Lawrence 1982, p1/4-5; Rose et al. 1989, p2) .

In addition to these requirements, careful consideration also needs to be given to procedures for emergency handling of passengers. Ideally, on new systems, a sidewalk should be located level with the vehicle floor, in order to provide a safe staff access and passenger evacuation route. Alternatively the ability to bring another train up to the rear of the one requiring assistance could be provided (Jeffries 1991, p203, Lawrence 1982, p1/6).

4.3.3.3 APPLICATION OF ATO

The inclusion of ATO enables the control system to be made to operate as a fully closed-loop, without the need for human interaction to complete the loop for either safety intervention or normal operation, as represented in Figure 4-7.

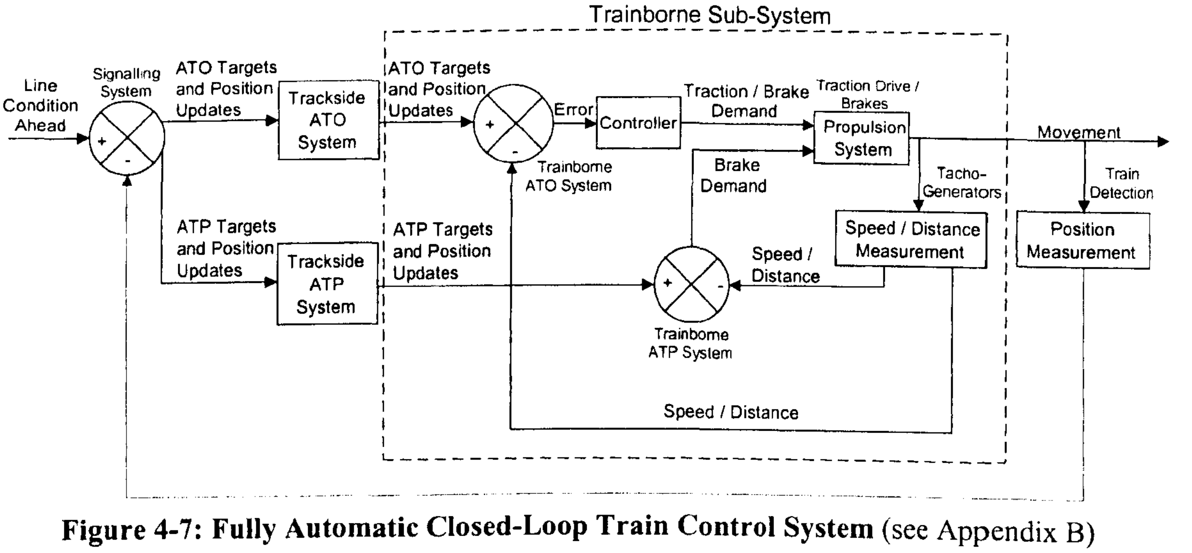


Figure 4-7: Fully Automatic Closed-Loop Train Control System (see Appendix B)

In order for an ATO system to operate effectively, the wayside sub-systems, trainborne subsystems and operating staff must all be integrated to form a cohesive whole.

Ideally, the whole line should be segregated and supervised from a central control room. Whether this is the case or not, the system must be designed with sufficient safeguards to ensure it remains as safe as (or safer than) existing manual systems of operation, where it is already the case that "drivers are unable to stop their trains before hitting staff on the track, passengers falling from platforms or obstructions across the track" (Lawrence 1982, p1/5).

ATO can be used to shorten journey times by ensuring maximum train performance. Alternatively, it can be used to reduce energy consumption significantly by use of coasting wherever the timetable permits. In either case, ATO can optimise line capacity by ensuring that all trains behave in the same manner (Goodman [(1)] 1999, p2). In addition to this, ATO can alleviate a major cause of train cancellations - staff absenteeism and sickness. It also avoids the problem of displacement of train crews following a disruption to the service (which can be a major factor in determining the rate of recovery of the service). Full automation also allows for quicker adjustment of the service to match demand, quicker recovery of service following disruption, quicker turnaround at terminals and the opportunity to continue full service off peak at marginal cost. As a result of these benefits, reviews of rapid transit administrations have shown that "significant automation of their networks has reduced operating costs, improved flexibility of operation and improved the reliability of services, particularly at off peak hours and weekends" (Lawrence 1982, p1/1 & 1/7). As an example of this, on Vancouver Sky train, "labour costs account for 57% of total operating costs, including all system maintenance and administration. This compares with a figure of 70-75% or more for a conventional transit system" (Jeffries 1991, p207). Some railway operators have also reported reduced wear on rolling stock (particularly instances of motor overheating and wheel flats) after the introduction of ATO (Milroy 1980, p2/11).

Despite the number and scope of benefits offered by the introduction of ATO, there are also major constraints on progress towards full automation. These include the social factors associated with reducing staff, the psychological factors in dehumanisation of the system and difficulties with policing the system to protect passengers from vandalism and physical attack (Lawrence 1982, 1/7). There are also operational problems within the concept of automated systems, the main one being the inability to recover from complete system failure and the associated risk of passengers becoming stranded without assistance being available (Dapre 1999, p13). Whilst ATO eliminates a number of undesirable attributes that human drivers introduce to the system, humans do have a remarkable ability to adapt to circumstances and, through rapid thinking and action, to avert danger and compensate for inadequacies in the design of systems (Stanley 1996, p2). Where ATO is introduced, it is therefore important to understand the roles that are played by both the technical system and a human driver. In the absence of the driver, it must then be ensured that the ATO system is more robust than the conventional systems it replaces, so that design inadequacies do not occur!

ATO systems are now in operation on a number of metro railways world wide (Eberhardt 2001, p3). However, "ATO has not been justified to date for main line railways" (Taskin et al 1995, pA3/14). The reasons for this are mainly that:

- Metro railways tend to operate at closer headways, with more frequent station stops than main line railways. They also tend to operate fleets of trains with roughly the same performance as each other. This means that the capacity benefits to be gained by ATO are proportionately higher on a metro railway;

- Metro railways tend to operate captive fleets of trains that are replaced at approximately the same time (usually in conjunction with resignalling work). This makes the installation and introduction of the new equipment easier to carry out, whilst also making it possible to change operating procedures and rules significantly at the time of introduction;
- Metro railways tend to be enclosed systems, making the problem of guideway intrusion hazards easier to manage.

Despite the historic reasons for not implementing ATO on main line railways, considerable benefits would accrue to them from its introduction. It has been noted that, as speeds increase and headways need to be reduced on main line railways, it will "persuade us to question whether we can leave the control of trains in human hands and still run on time" (Rayers 1989, p14). When it is further considered that the technology to achieve more robust systems of guideway intrusion detection is now available (or at least practicably conceivable) and that the use of other ATO technology has been proven in service on metro railways, it appears likely that the use of ATO will extend to main lines in the near future. Indeed, operational tests of ATO have been conducted on the S-Bahn near Dresden in Germany.

4.3.4 AUTOMATIC TRAIN SUPERVISION (ATS)

Whereas the closed-loop control systems considered so far have been from the drivers' perspective and the automation systems considered have been intended to overcome the problems associated with a human driver, there is a second large human element to the conventional operation of a railway. That is the operation of the signalling system by signalmen. Automatic Train Supervision takes on the automation of the signallers' and controllers' roles. It is therefore responsible for the monitoring and co-ordination of individual train movements in line with the schedule and route assignments (Khessib 1989, p10). Hence a definition can be given as:

> **Automatic Train Supervision (ATS): -** 'a generic term for describing systems which reduce the reliance on human involvement in the central control functions of a railway'.

As shown in Figure 4-5, ATS incorporates the functionality of ATR, ARS and Automatic Traffic Monitoring / Service Control. ATS and its constituent parts may be present in the system with or without ATC and, where fully implemented, would include:

a. Automatic tracking of the location of all trains, including monitoring of track circuit information, point information, train identification, etc.;
b. Provision of an interface with staff at the control office, including equipment states and train locations;
c. Fleet and Staff management;
d. Attempting to keep trains running to a pre-defined timetable or even service interval by control of dwell time and use of coasting;
e. Deciding precedence of trains at junctions, resolving conflicts according to a pre-defined algorithm or priorities;
f. Automatic control of route management and route setting;

g. Automatic development and implementation of strategies to restore planned service following disruption (turning back and cancelling trains, use of crossovers and bi-directional facilities, etc.);
h. Sharing of information regarding conditions between all affected trains (e.g. modifications to speed profiles to allow for localised slip / slide problems, appropriate use of regenerative braking, etc.);
i. Logging of all operations for legal and commercial purposes;
j. Simulators (for training and to help replay incidents).

(Dapre 1999, p16; Mott 1993, 3.3.4/3; Appendix A Interviews)

ATS is a non-vital system, the failure of which would not directly affect railway safety.

4.3.5 AUTOMATIC TRAIN REGULATION (ATR)

As outlined in section 4.3.4, Automatic Train Regulation (ATR) is a sub-set of ATS. The purpose of regulation is to:

a. Keep train running to a pre-defined timetable or even service interval, by control of dwell time and use of coasting;
b. Decide precedence of trains at junctions, resolving conflicts according to a pre-defined algorithm or priorities;
c. Automatic control of route management and route setting;

(Appendix A Interviews)

Therefore, ATR can be defined as:

> **Automatic Train Regulation (ATR):** - 'a system to ensure that the train service maintains schedules or, following disruption, returns the service to timetabled operation or to regular fixed headways, by automatically adjusting the performance of individual trains'.

Regulation to maintain an even service interval has in the past been performed very crudely through the signalling system by delaying individual trains in order to prevent long gaps arising in the service or to provide compliance with the timetable. "Before the advent of centralised train control (CTC), the function of regulating the movements of trains was carried out locally by station masters, negotiating with neighbouring colleagues. Increased and faster traffic required an overview and train regulation therefore became the task of dispatchers" (Schmid 1999, p3). These traditional approaches to regulation had the disadvantages of delaying customers who would prefer to be carried to their destination as quickly as possible, and of causing compound delays that must eventually result in a train cancellation or turn back in order to recover normal service patterns.

With the advent of more sophisticated modern control systems, alternative regulation techniques are now possible. This typically involves running the normal service with coasting, rather than using the train's maximum performance, to provide a number of regulation benefits:

a. where perturbations occur to the service, an even service interval can be maintained by speeding trains up as well as by slowing them down;
b. by regulating the speed of trains on the approach to junctions so that they do not arrive until the route is available for them, the need to bring trains to rest is avoided, hence giving run-time improvements;

c. the intelligent use of coasting results in energy efficiency improvements. (McCormick, et al. 1996, A1/13).

This approach requires regular reviews and adjustments to train progress. Where trains are manually driven, some advantage can be obtained by passing advice to drivers on the optimal speed profile ahead of their train, but this technique is really ideally suited for application to automatically driven trains.

As with ATS, ATR is a non-vital system, the failure of which would not directly affect the safety of the railway.

4.3.6 AUTOMATIC ROUTE SETTING (ARS)

A subset of the functionality included within ATR is commonly known as Automatic Route Setting (ARS). Perhaps the simplest form of ARS is Automatic Block Signalling (ABS), which has already been outlined in section 3.8. Under ABS there is only one possible route from a signal and the aspect displayed by the signal is determined, without intervention, based upon the state of axle counters or track circuits and certain other conditions (e.g., level crossing closure). However, the term ARS is usually reserved to describe a system of automation where selection between possible routes is required, which is not the case for ABS. Therefore, ARS can be defined as:

> **Automatic Route Setting:-** 'an electronic or relay based system which sets routes using information from a train describer and the timetable without the need for intervention by a signaller.'

The train describer system manages all train numbers and tracks the position of each train as precisely as possible, based on information from the interlocking (Kuhn 1998, p16).

In its simplest form, this definition of ARS describes the type of system referred to on LUL as programme machines (State of the Art 1998, p342). This system of operation utilises roles of punched plastic film that contain details of the expected trains and their required routing or timing, in accordance with the timetable. Machines of each type can be coupled together to ensure that, as long as trains arrive in the assigned order, the required routes are automatically set at the correct times. If a train is detected for which the train describer information does not agree with the punch card programme, or a train does not arrive when expected, an alarm is sounded in the control centre. The control centres are equipped with a switch for each programme machine, enabling selection of route setting by programme machine, on an automatic 'first come, first served' basis or by the signaller. There is also a train cancelled push button to step the machine to the next entry if required (Dell 1958, p84,100). In case of serious disruptions, the punched film can be replaced by an alternative timetable.

On main lines in the UK, the term ARS is applied to a more complex system than this, which also includes some basic regulation functions at a localised level. The prototype of the BR ARS system was commissioned at Three Bridges on the Southern Region in 1983. The system receives information from the train describer system and the Master Timetable System, and uses this to perform route setting for all time tabled train movements (even when the service is disrupted). Many of the other routine activities of the signalman, such as train monitoring, track circuit monitoring, automatic code insertion and timetable handling are also performed by ARS. The signalman can then focus on more serious problems, such as stock failures and signalling failures. Because of the scope of BR ARS, It has been argued (and the author supports the view) that it should more accurately be described as a form of ATR or even ATS, since it does more than just set routes (Hurley 1991, p334-5; State of the Art 1998, p341-2).

The author's definition of ARS is in line with the route setting functionality of BR ARS. This functionality of the system continuously monitors all trains and determines which require routes to be set. The general principle is to keep two signals ahead of the train at Green. Where a station stop is scheduled, the ARS only sets the route ahead once the 'train ready to start' signal is received from platform staff or a pre-set time before the scheduled departure. When ARS determines that it should set a route for a train, it checks whether the route is available. If it is, a request is sent to the interlocking. If not, ARS can register a flag to reconsider the train once the route becomes available or can select a warning or calling on route instead. If ARS is unable to perform its route setting function, it generates an alarm for the signalman. This occurs for:

- any train approaching the area of control that can not be found in the timetable;
- if the signalman routes a train off of the timetabled path ;
- after 4 attempts at setting a route without it beginning to set;
- if a route does not complete setting within a specified time (usually indicating a point detection failure)

(Hurley 1991, p335-6; Leach 1991, p218; State of the Art 1998, pp341-2).

The facility for local amendments to the timetable (such as addition of an un-time tabled train) is also provided. This facility includes some default patterns that the signalman can select and assign to an untimetabled train, so that BR ARS can then operate for that train. This facility can also be used to replace the timetable with a contingency timetable designed to recover from a train or signalling failure (Hurley 1991, p337).

As previously mentioned, the BR ARS also performs basic regulation activities through functionality for conflict resolution. Whenever an extra section of route needs to be set for a train, the system checks that the request does not conflict with any other planned train movements. Where a conflict is detected, the system predicts the future running of each train. If either train can continue without delaying the other, it is allowed to do so. Otherwise the train that would cause the least delay (weighted by a factor of the train class, route required and time of day) is allowed to proceed first (Hurley 1991, p335; Leach 1991, p215-8).

In addition, the BR ARS performs some traffic monitoring and service control functions:

- Train Monitoring – when a train passes a signal, the time at which it should pass the next one is predicted. If this prediction is exceeded by a specified margin when the next signal is not red, or just cleared from red, a 'train unusually long in section' alarm is raised;
- Track Circuit Monitoring – correct sequence of track circuit operation is monitored as trains pass through the area of control. When a track circuit becomes occupied or clears unexpectedly, an alarm is raised to the signaller;
- Automatic Code Insertion – when a train reaches a station where it is time tabled to split or to terminate and turn back, the system can automatically insert the required train code into the train describer system;

(Hurley 1991, p335-7; Leach 1991, p219; State of the Art 1998, pp342).

The main attraction of ARS systems is a significant reduction in the signallers' workload. At Liverpool street, for example, introduction of the BR ARS system allowed a reduction from three to two signalmen and decreased the workload of the remaining two (Hurley 1991, p342). The success of ARS does, however, vary between signal boxes. The greatest benefits for the BR ARS have been obtained where there is a high degree of consistency in the service pattern. "Where there is greater diversity of traffic it is a little less perfect and calls more on the ingenuity of the signalman to maintain the optimum service" (Francis 1994, p5). Care must also be taken in developing the route setting strategy for the ARS, since setting too early or too late can result in delays to the service, either in normal operation or when a route fails to be established at the initial attempt (Francis 1994, p5).

4.3.7 AUTOMATIC TRAFFIC MONITORING / SERVICE CONTROL

The traffic-monitoring element of ATS would generally be responsible for supervising traffic movement. This would include monitoring of track circuit occupations and the time taken by trains to pass through each section (as described for BR ARS in section 4.3.6), along with recording of information for future analysis. It would also include watching for serious perturbations to the service from which regulation alone is unlikely to recover. Where serious perturbations are observed, action may need to be taken such as turning trains back to fill in the service, cancelling trains and other higher-level decisions about divergence from the time tabled service pattern (Taskin et al. 1995, pA3/15; Appendix A, Interview with Jerry Lewis).

The activities associated with recovery from serious perturbations to the service include management of the service in real time to ensure that train destinations are appropriately balanced, that bunching / conflicts are minimised, and that staff and stock resources are available when and where required. This functionality was defined as Service Control in section 2.6.2.

The author is not aware of any system currently in use that automatically carries out service control, probably due to the complexity involved in deciding on the right course of action to recover from serious service irregularities. It would, instead, be the usual practice for the automated system to operate an alarm to draw a human operator's attention to the need for action and, subsequently to provide information to support decisions by that operator. However, as technology continues to advance, automation of the complete response is becoming more feasible.

4.3.8 HUMAN FACTORS IN AUTOMATION

The motivation for introducing automated systems is usually to reduce workloads (and thus staffing levels) and to improve on the safety performance of human operators. "A common response to the complexity of human information processing is to suggest that by automating the entire system human frailty can be eliminated" (Moray 2001, p15). However, even highly automated systems still require human intervention at times (IRSE[(1)]2001 Paragraph 50) and "although it might seem that automation would decrease the risk of operator error, the truth is that automation does not remove people from the system – it merely moves them to maintenance and repair functions and to higher-level supervisory control and decision making. The effects of human decisions and actions can then be extremely serious" (Leveson 1995, p10). In the discussions of automation systems so far, it is mainly the advantages and method of operation that have been considered. In this section, the author therefore identifies a few areas of consideration that should be given to the human factors implications of automation – both during normal operation and under degraded modes.

The potential for human error to contribute to train accidents is significantly reduced by the introduction of automation systems, such as ATP, but is not eliminated altogether. Issues still existing include:

- **Errors in Preparation:** Errors in system design or the data on which the automated systems act may be subtle and not reveal themselves for a long time, or may be introduced as a result of wrongly executed actions and misrepresentation during data entry by an operator (Moray 2001, p16; Short 2001, p4);
- **Maintenance Errors:** Automated systems must be maintained by humans, who are often less well qualified than the operators they have replaced (Moray 2001, p16);
- **Isolation from System:** "Operators in automated systems are often relegated to central control rooms, where they must rely on indirect information about the system state: this information can be misleading" (Leveson 1995, p10);
- **Delayed Reactions:** If the automation is extremely reliable, human supervisors may become 'complacent' and cease to monitor the state of the system. They may then be unable to react or take effective control in a timely manner following failures in that system (Francis 1994, p5; Kecklund et al. 2001, 9; Moray 2001, p16);
- **De-skilling:** De-skilling can be a major cause of inappropriate action being taken. As an example of this, the use of an ATO system may lead to reduced driver ability to drive trains manually in a safe manner following equipment failure (Smith 1999, p2);
- **Communication Errors:** Errors in verbal communication are most critical in relation to degraded modes of operation. Most forms of ATP, for example, will do nothing to prevent this, as they generally include manual override facilities (Short 2001, pp3-4);
- **Attention Conflicts:** Conflicts, such as requiring a driver to look at both in-cab displays and lineside signals, may cause important information or events to be overlooked (Kecklund et al. 2001, p6);
- **Incorrect Assumptions:** If not all trains on the network are fitted with ATP or there has been a failure of the ATP system, staff assumptions that the automated system is providing for safe operation will not be correct and may lead to inappropriate actions. (Smith 1999, p3).

When considering the above points, it is worth noting that engineers can only automate what they understand. Hence, if automation is taken to its limits, it is those aspects of the system that are too difficult for engineers to automate that are left to the human operators (Bainbridge 1983). This in turn means that the remaining tasks are more complicated for the operator and have to be carried out with a reduced understanding of what is happening within the system – making the occurrence of human error all the more probable. In addition to this, if the automation of functions within a system makes the remaining human operators subordinate to the machine, rather than the machine subordinate to them, they may loose their sense of purpose, resulting in poor quality of work or high employee turnover (Rosenbrock 1990, pp114-5).

Since the use of automation does not eliminate the opportunity for, or effects of, human error (and may even make the situation worse for some types of error), it is important to consider the role that humans play within the system and to design the system accordingly. In other words, an overall systems approach must be taken, considering the effects of technical failures and interactions with staff, passengers and the public (Smith 1999, p8). This is an approach fundamental to the management of system safety, as highlighted by the Hazards Forum:

> "The human beings who operate safety-related systems, and those responsible for the use and operation of such systems, whether in a supervisory capacity or directly manipulating controls, must also be regarded as part of those systems" (Hazards Forum 1995, p3)

Based on considerations such as these, "the consensus is now that a combination of human and machine is better than either alone" (Moray 2001, p15). In general, "computers are better at drawing simple conclusions from large amounts of data (deduction) whereas humans are better at drawing complex decisions from small amounts of data (induction). An information system which can take care of deductive reasoning and provide the operator with reliable, salient information from which he can perform inductive reasoning is most likely to optimise operator performance" (Cobb et al. 1996, pp10-11). It has further been noted "humans have a remarkable ability to adapt to the circumstances and to avert danger through rapid thinking and action, sustaining process reliability by human intervention and compensating for inadequacies in design" (Stanley 1996, p2). In accordance with this approach, machines should be designed to assist, rather than replace, the skills and abilities of human operators (Rosenbrock 1990, p166). Thus, rather than striving to minimise the human involvement in the system, consideration should be given to the role and work load of the human components of the system, such that the risks that they introduce are minimised, whilst maximum use is made of the strengths that they can contribute. "It is very clear that simply 'engineering out the human' is not a wise solution" (Moray 2001, p17).

4.4 AUTOMATIC TRAIN PROTECTION, SYSTEMS AND APPROACHES

In Appendix C, the author has outlined what could be described as a plethora of ATP systems that have been developed over the years. The outlines include discussion of the relative merits and limitations of the different systems and show that there are lessons to be learnt from all of them. The lessons learnt from producing this appendix were fundamental to the direction taken during the study documented in this thesis. The author has, therefore, included this section within the main thesis in order to summarise the findings of Appendix C and explain the focus of the remaining chapters in this thesis. He would also highly recommend that the reader study the appendix itself in detail.

Several common themes can be drawn from Appendix C that are particularly worthy of note when considering the future optimisation of train protection systems. These are:

- A number of systems have suffered from high costs in proportion to benefits for design, installation and maintenance, often making them un-viable. e.g. the NER Electric Cab Signalling (section C3.4), SRAWS (Section C4.4), BR ATP (section C6.1), TVM 300 (section C7.2), TVM 430 (section C7.3) and, LZB (section C7.4);
- A number of systems have suffered from poor reliability, often making them un-viable. e.g. the Great Central ATC (section C3.5), the Great Central Reliostop (section C4.3), BR ATP (section C6.1), SELTRACK (section C7.7);
- A number of systems have restricted, or are being predicted to restrict, the achievable headway. e.g. Mechanical Trainstops (section C4.1) and ETCS level 1 and 2 systems (sections C6.7 and C7.10);

- A number of systems have given rise to safety concerns where human factors have not been eliminated from the safety loop at all times. e.g. all of the warning systems (section C2), Automatic Warning Systems (section C3) and Partial ATP systems (section C5).

If ATP implementation is to be optimised, these limitations must be considered and resolved in future systems.

A number of the historical systems trialled in the UK failed to achieve success due to weaknesses or expense in available technology, rather than fundamental failures in their operating principles. It is to be regretted that later generations of engineers did not resurrect some of these systems (which offered far more comprehensive protection than many of their successor systems) once technological developments had overcome their original limitations. If they had done so, today's railways might have been made considerably safer. However, that is now in the past (although a salutary lesson none-the-less) and we must look towards the future.

The development of standardised systems (such as the ETCS level 1, 2 and 3 systems) should go a long way towards overcoming the key limitations that have applied to previous systems:

- All three levels of ETCS will all offer comprehensive ATP, reducing the reliance on human factors in ensuring the safety of train movements;
- They should reduce overall system costs by enabling a higher rate of manufacture (through increased market size) and encouraging competition between suppliers;
- It is also to be hoped that standardisation and competition will give rise to more robust and reliable equipment than has been available for some of earlier systems.

This leaves one main area of concern for future development of optimised ATP systems – the effects that the ATP has on achievable capacity.

4.5 SUMMARY

One of the main causes of accidents is human error. There is therefore the possibility of increasing railway safety by considering human factors within system design, in order to find ways to reduce the incidence of those errors. However, recent studies have suggested that little benefit can be expected from improvements in human performance alone. Instead, the most promising improvements appear to be offered by the potential to automate parts of the system and thereby remove the causes or mitigate the effects of human errors.

Recent advances in technology have seen the development of secure mechanisms for the transmission of data without fixed connections, combined with the advent of sophisticated and reliable systems for the processing of that data. These have in turn provided the required medium for implementation of automatic train control and supervision systems, whilst opening up the potential for the replacement of fixed wayside signals with in-cab signals and for the implementation of moving block control.

In this chapter, the author has discussed the nature of transmission based signalling and the possibilities for automation in train control systems. In so doing, he has offered definitions of the terminology associated with these types of system. He has also highlighted the continuing need for the consideration of human factors as an integral part of the design of automation systems.

In chapter 2, the purpose of train control systems was identified as being to ensure that trains are enabled to run safely and efficiently. It follows from this that the impetus behind the development of advanced methods of train control has also been safety and efficiency. Of all the developments considered within this chapter, that with the greatest potential to influence safety is Automatic Train Protection. Unfortunately, this has also been identified as a potential cause of reduction in railway capacity (which equates to a negative effect on efficiency). The author, therefore, considered some of the approaches that have historically been implemented to achieve ATP in more detail (as outlined in Appendix C and summarised in section 4.4).

Based on the background research conducted in producing chapters 1 to 4 and appendices A to C, the author concluded that the main area of concern for future development of optimised ATP systems is the effects that those systems will have on achievable capacity. This conclusion led the author to identify a number of areas for further study:

- Modelling of the relative capacity impacts to be expected from different ATP system types, in order to assess which of the systems would be most suitable for future use – and produce quantitative measures of system capacity in support of that assessment;
- Consideration of other railway system factors, beyond the direct influence of ATP system design, that might also contribute to overall capacity (in particular the relationship between train braking performance, speed restrictions and ATP

5 TRAIN BRAKING PERFORMANCE

5.1 INTRODUCTION

"In railway operation, it is not how fast a train can run which limits performance, but how well, and consistently, it can stop" (Harris et al. 1992, p119). Indeed, "braking is the most important control function in transport because it determines the capacity of a traffic lane" (Barwell 1983, p190).

All comprehensive ATP systems rely on the use of the trains' service and / or emergency braking systems to bring them under control if the supervision criteria are infringed. The brake rates achievable by the train (including any brake application delays) are required to determine the ATP system's targets and supervision criteria. The rates must be known either in the design process (for a speed code type system) or as data for use in the real-time calculation of braking curves (for a distance to go type system). This means that the trains' braking performance is a central factor in determining the efficiency with which the ATP system will operate – and hence the capacity that will be achievable on the railway. The assumed braking characteristics must be conservative enough to ensure protection under all reasonably practicable conditions, but not so conservative as to unduly constrain operations. As a consequence of this, a thorough understanding of train braking is fundamental to the optimisation of ATP systems.

In the debate following an IRSE lecture in 1949, W H Challis noted that, if braking performance could be significantly improved, the problems of signalling for high speeds would be largely overcome (Nock 1949, p158). In the 2001 IRSE presidential address, it was noted that train speeds and braking distances impact railway capacity far more than imperfections in signalling and train control systems. The president then concluded that "we need to look at radical options for our railways and transit systems in the future, if these better match society's expectations" (Barnard 2001, p5). This suggestion was echoed in the findings of the Southall and Ladbroke Grove joint enquiry, which recommended "research into other means of stopping trains, including enhanced emergency braking (EEB), should be pursued vigorously" (Uff et al. 2001, p xii, item 10). It is, therefore, not enough to determine the potential for optimising the representation of current braking systems within ATP systems. If the capacity of the railway system as a whole is to be optimised, the potential for new approaches to train braking must also be considered.

Linked with the subject of train braking performance is that of the wheel/rail interface. In a report to the UK government in 2000, Sir D Davies highlighted the topic of wheel/rail interface and adhesion in particular, as underpinning the entire operation of railway systems. It was noted within this report that poor adhesion can cause "very dangerous situations that are not ATP preventable and we need to understand how they arise and find ways to minimise their effects" (Davies 2000, p44).

In this chapter the author considers the subjects of train braking and the wheel/rail interface. The intention is to develop an understanding of the issues related to train braking performance that should be considered when implementing train control systems and also to determine the potential for improving train-braking performance in the future.

5.2 THE DYNAMICS OF TRAIN BRAKING

When a train is accelerated, the work done is stored as kinetic energy. When it is moved up an incline, the work done is stored as potential energy. As the train moves along the railway its kinetic and potential energies will vary. However, since energy can not be created or destroyed, the sum of its kinetic and potential energies will remain the same unless work is done (energy is added or removed) by an external force. When such work is done, the sum of the train's kinetic and potential energies will change in direct proportion to the work that has been done. This energy balance is represented in equation 6.1.

$$EK_S + EP_S = EK_F + EP_F + W \quad \text{Equation 5-1}$$

Where:

- EK_S: Kinetic energy of the train at its start location
- EP_S: Potential energy of the train at its start location
- EK_F: Kinetic energy of the train at its final location
- EP_F: Potential energy of the train at its final location
- W: Work of the braking and tractive forces

(Based on Monselet 2001p, 97)

Changes in a train's energy balance may come about for a number of reasons:

- The train's traction system converts the energy stored in hydro-carbon fuels or electricity supplies into tractive effort (acting as an 'external force' to increase the energy balance);
- The natural effects of drag and friction forces convert the train's energy into heat in the surrounding environment (acting as 'external forces' to decrease the energy balance);
- Forces applied by the train's braking systems convert the train's energy into another form, typically electricity or heat (acting as an 'external force' to decrease the energy balance);
- The gradient of the track (the angle from the horizontal) results in the altitude of the train changing as it travels along. This causes the train's kinetic energy to be converted into potential energy on an incline and vice versa on a decline.

In accordance with the energy balance, a train's braking systems must be able to carry out enough work to convert the train's kinetic energy into other forms at an adequate rate to reduce the speed to a desired level within the desired distance. This reduction in speed must be achievable even if the train is on a gradient, such that its potential energy is being converted into kinetic energy as it moves along. Also, as the work done will result in conversion of the train's kinetic energy into heat or electricity, the braking system must also be capable of dissipating the resultant heat or electricity without damaging itself or its environment.

Since the kinetic energy of a moving body increases with speed, "higher speeds produce longer braking distances and greater headways, unless braking is improved" (Broadbent 1969, p2).

5.3 BRAKING MODES

Train braking systems are required to operate in three ways:

1) Service brake - providing accurate stopping and preventing rollaway in service, without causing an undue degree of acceleration and associated jerk to the passengers at any time;
2) Emergency brake - stopping the train in the shortest distance possible without causing injury to passengers;
3) Parking brake- ensuring that a stationary train remains so when not in use.

In an emergency the shortest possible braking distance is required. Therefore, deceleration rates must be as high as possible and application delays as short as possible. In the UK, requirements for emergency braking are enshrined in the regulation of railways act 1889. This states that passenger trains must have continuous brakes that apply automatically in the event of "any failure in the continuity of its action". That is, if the train divides. The same act also requires the brake to be in regular use in daily working (Broadbent 1969, p76).

Different railway authorities have differing requirements for emergency brake performance. On Network Rail lines, the legal requirements of the regulation of railways act 1889 are repeated in standards that require the emergency brake to apply automatically if control signals or power sources are lost, or if the train divides. On LUL they are met by requirements that only permit emergency braking to utilise friction brakes (see section 5.5) with a fail-safe application method (Fargher[1] 2000, pp8-9; LUL[1] 1991, pp3-4).

The meaning implied by the term 'emergency brake' is not the same in all countries. As an example of this, the UNISIG ERTMS System Requirement Specification considers an emergency brake application to be one that can be relied upon to ensure safety (UNISIG 2001, p3/62). This treatment allows for a very different emergency brake arrangement. Where UK railway operators utilise only fail-safe friction braking techniques (see section 5.5) for an emergency brake application, non-fail-safe dynamic, eddy current and magnetic track brakes (see sections 5.5.2 and 5.5.3) have also been used for emergency braking on DB in Germany and the Shinkansen railway in Japan (Saumweber 1986, p2; Kumagai 1996, p224), with safety being ensured by limiting the proportion of the expected brake force that could be affected by any single failure within the brake system.

The requirements for service braking differ somewhat from the requirements in an emergency. The operating department will be concerned with running times and passenger comfort, making demands for "the braking equipment to produce stops with passenger trains in which the braking time is the shortest possible for the service concerned, consistent with an absence of any degree of shock affecting passengers unduly at any time during the stop" (Broadbent 1969, p2). Here, safety is not the primary consideration since the emergency brake can be used as a fall back if required. Therefore, the service brake need not be fail-safe and, in accordance with this view, the UNISIG ERTMS System Requirement Specification refers to a service brake application as one that "is considered not safe" (UNISIG 2001, p3/62).

Many train braking systems utilise the same components for service and emergency braking, the only differences being the method of actuation (which will be more assured and possibly quicker for the emergency brake). This situation is represented in Figure 5-1. However, in other braking systems a higher instantaneous brake rate may also be achieved in emergency braking, either by applying higher pressures to existing frictional elements or by introducing an entirely separate braking mechanism.

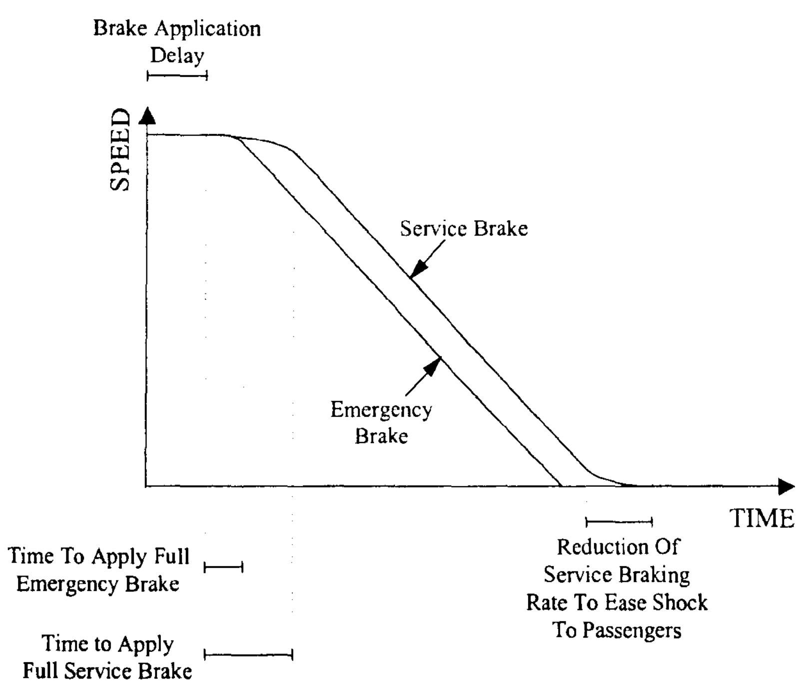


Figure 5-1: Comparison of Service and Emergency Brake Application (Broadbent 1969, p3)

Parking brakes are intended to hold, rather than stop a train. Railtrack defined parking brakes as "a brake system designed to hold a rail vehicle stationary for an indefinite period without the addition of further energy to maintain the brake force, provided no additional external force is applied to the vehicle" (Fargher(1) 2000, p5).

Parking brakes almost exclusively use conventional friction brakes (tread or disk), although permanent magnet track brakes can also be used. Usually, the same braking components as the service brake are used, being applied by a different mechanism to ensure they remain applied without brake fade effects. However, some freight vehicles have a parking brake that is wholly independent of the service brake system.

Typically the method of application for a parking brake would be one of:

- Spring applied / air released (which is fail-safe, loss of the train's air supply causing it to apply);
- Hydraulic application and release;
- Mechanical operation (by operating a wheel or lever that mechanically applies a force to the normal brake rigging);

  (see appendix A, interview with K Schofield).

Where trains are formed of carriages that may be left on their own, as is common for freight vehicles in Britain and most trains in continental Europe, individual parking brakes are required on each carriage. However, parking brakes are expensive and heavy and it is undesirable to over specify them (Schofield[1] 2000, p9/7). Therefore, where trains are operated in standard formations, fewer parking brakes are fitted. Typically in the UK, a multiple unit will have sufficient parking brakes to hold itself on a 1 in 40 incline, a locomotive for 1 in 30 and a passenger train for 1 in 100 (see appendix A, interview with K Schofield).

5.4 BRAKING LIMITATIONS

A number of limitations apply to the braking rates that trains can achieve in practice. The most obvious of these is the rate at which the train's braking systems are capable of converting its kinetic energy into other forms (such as heat or electricity). This depends on the brake mechanism and materials used, some of the options for which will be discussed in sections 5.5 and 5.6. However, with the equipment and materials available for use on modern trains, the energy dissipation performance that can be achieved is not, generally, a serious limiting factor to braking rates. A far more significant factor, and therefore the first that will be considered, is the coefficient of adhesion that exists between a train's wheels and the rails on which they run.

Also of interest is the stress that can be safely or comfortably withstood by passengers on the trains. That is the effect of acceleration and jerk rates, which will also be considered.

5.4.1 ADHESION

The coefficient of rail / wheel adhesion (sometimes called the coefficient of friction) is the ratio of the weight of the vehicle to the force that can be used for traction and braking. It is either expressed as the pure ratio (μ) or as a percentage.

5.4.1.1 OVERVIEW

Under laboratory conditions, steel on steel has a coefficient of friction of 0.7-0.8 (70-80%) (Technology Primer [2] 1997, p97). Unfortunately, in the real world, the wheel/rail interface is subject to a wide range of contamination, such as water (including rain, snow, frost, ice and dew), mud (particularly at level crossings), rust, oil, grease, leaves, insect infestation, industrial contaminants and detritus from brake friction materials (Broadbent 1969, p11; Gibson 1996, p26; Nagase 1989, p39; Sergeant 1996, p87). Some of these can be controlled, whilst others, such as those caused by environmental conditions, are uncontrollable (Monselet 2001, p68).

The level of adhesion achieved also depends in part on the rail condition and profile. Track irregularities such as undulations, corrugations and curvature can cause sudden weight transfer from wheel to wheel, with resulting "loss or reduction in contact force, and probable loss of adhesion" (Sergeant 1996, p88).

The effect of contaminants and irregularities such as these is to reduce the adhesion levels likely to be achieved in practice to well below 70%. As an example of this, the relationship between speed, water contamination and adhesion can be seen in Figure 5-2.

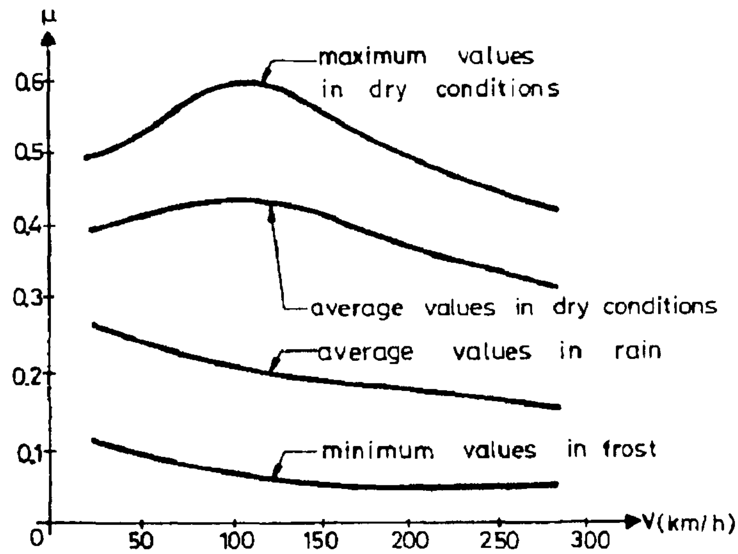


Figure 5-2: The Relationship between Adhesion Coefficient (μ), Train Speed (V) and Climatic Conditions (Profillidis, 2000, p228)

When a vehicle relies upon the wheel/rail interface to achieve braking, the coefficient of wheel/rail adhesion acts as a limiting factor. "If an attempt to brake at high rates is made it is likely that on some occasions the axle will decelerate at a greater rate than the vehicle and wheel" (Watson 1999, pM07/1). This occurrence is refereed to as 'slide'. The equivalent effect during acceleration is referred to as 'slip' and the combined effects of both slip and slide are referred to as 'creep'. "Creep is the fractional difference between the peripheral speed of the wheel and the speed of the train" (Abuzeid 1996, p2).

At a slide rate of 100%, the wheel ceases turning and begins to slide along the rail. This can bring about damage to the wheel tread surfaces and may also increase the stopping distance, since the sliding wheel / rail adhesion coefficient is generally lower than the non-sliding coefficient (Watson 1999, pM07/1). However, studies have shown that at low levels of creep, the adhesion coefficient actually increases. An example of this can be seen in Figure 5-3, which is based on measurements taken by General Motors. The figure shows the effects of creep on adhesion under a range of environmental conditions.

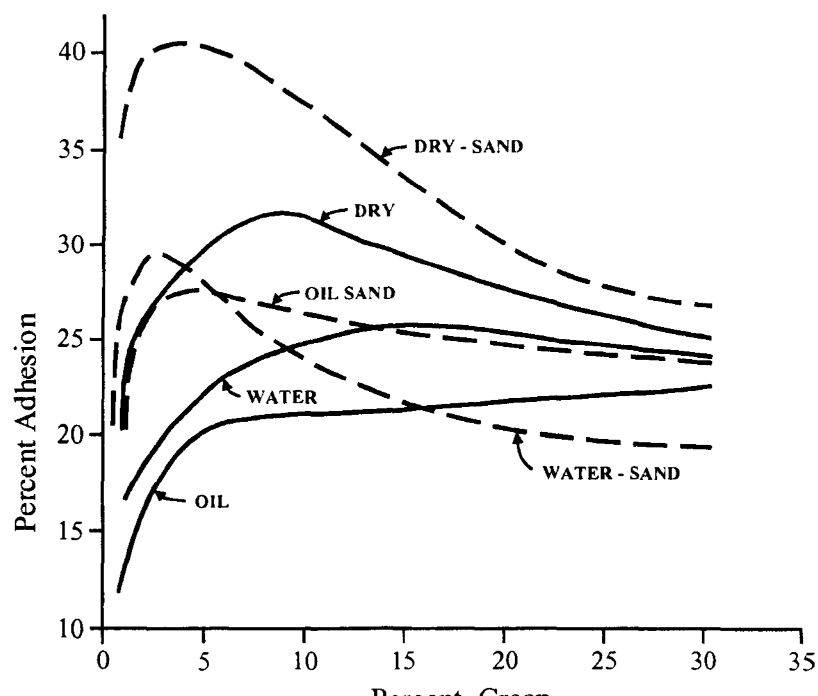


Figure 5-3: Effects of Creep and Rail Condition (Technology Primer [2] 1997, p97)

The effect of 'creep' varies with environmental conditions and speed, but use of adhesion is usually maximised when the peripheral speed of the wheel is 3 to 10% slower or faster than the speed of the train (Nagase 1989, p41). This improvement in adhesion is in part due to a cleaning effect produced when the wheel in 'creep' polishes the track. This effect not only improves the adhesion achieved by the wheel in creep, but also that achieved by following wheels which benefit from the cleaner track (Schofield[1] 2000, p9/26).

Testing by British Rail showed that the adhesion of a wheel in creep peaked at about 5% creep, initially falling off sharply as the rate increases, but stabilising between about 15 and 25%, before then falling at a lower (and approximately linear) rate until wheel lock up occurs at 100% creep. The same tests also found that the wheel/rail conditioning properties increase as creep becomes higher, thus improving adhesion for following wheels (Fulford 1997, p25). Unfortunately, excessive levels of creep can cause metallurgical damage to the rail. A sophisticated and stable control system is required, therefore, if the effects of creep are to be optimally harnessed (Gostling 1986, p4). In order to do this, WSP systems (see section 5.4.1.2) are generally designed to use the fairly stable characteristics of 15 to 25% wheel creep (Fulford 1997, p25).

In dry climates and on underground railways, adhesion levels of 0.4 to 0.6 are regularly encountered, with a figure of around 0.14 usually considered to be the maximum that can be relied upon (Broadbent 1969, p13;Tunley 1999, pM13/1). However, during the leaf fall season in a wet and open environment (such at the UK main lines), adhesion levels "between 0.05 and 0.09 are quite common, and very low adhesion conditions (below 0.05) occur frequently enough to cause significant operational and safety problems to the railway" (Tunley 1999, ppM13/1). In extreme cases, leaf film can reduce adhesion to as low as 0.01 (Shooter 1993, preface).

The levels of adhesion assumed by railway operators vary. On London Underground, the target railhead friction level is 0.25 to 0.35 (Mansfield 1998, p14). Research on the Central Line has identified the probabilities of low adhesion during brake applications over 100m of track to be 0.33% for μ below 0.1, falling to 0.0016% for μ below 0.03. On the basis of these findings, overlaps on the Central Line have been validated for average adhesion down to 0.03 over 100m lengths and 0.06 over 200m lengths (Chandler 2001, p27). In order to minimise the risk of wheel damage in emergency braking, the standard approach on LUL is to assume a nominal retardation rate of 1.1m/s^2 in the open and 1.3m/s^2 in tunnel areas, with the instantaneous retardation being limited to 1.8m/s^2 at speeds above 20km/h and to 1.98m/s^2 at lower speeds (LUL[1] 1991, p5; Rowe 1993, p31). In contrast to the Central Line, two series of friction measurement on LUL's Northern Line revealed actual friction to be 0.3 on straight track railhead and 0.29 on curved track railhead, indicating that higher braking rates could be supported (Mansfield 1998, p14).

Since 1964, the Shinkansen railway in Japan has used an equation for the assumed lower limit of adhesion coefficients under wet rail conditions:

$$\mu = \frac{13.6}{(V+85)}$$, where V is the train speed in km/h. Equation 5-2

This equation (which assumes adhesion of 0.16 at rest, falling to 0.1 at 51km/h and 0.06 at 130km/h) has been assessed against experimental results on a roller rig and test running on conventional narrow gauge lines. It is used to limit the brake force applied on Shinkansen trains as their speed increases (Kumagai et al. 1996, p220; Ohyama 1991, p19). More recent test results on a number of lines in Japan have suggested, however, that the adhesion experienced in practice is likely to be quite different. In dry rail conditions, travelling at speeds of up to 50km/h, adhesion levels were generally much higher than predicted by the equation (averaging over 0.2 and not falling below 0.15), with hardly any link being found between the train speed and measured adhesion. After rainfall, when travelling at speeds of up to 80km/h, adhesion levels were still higher than predicted. Average adhesion was found to exceed 0.2, falling below 0.15 only at level crossings (where mud had been deposited on the rails, reducing adhesion to near 0.1). However, at the beginning of a period of rainfall (when travelling at up to 70km/h) and during frosty conditions (travelling at up to 40km/h), adhesion levels averaged nearer 0.05, falling to near zero in areas of leaf fall (Nagase 1989, pp35-7, 41).

On main line railways in the UK, signalling systems have traditionally been designed to assume train braking rates of 7%g (achieved by use of tread brakes) for operation up to 160km/h and 9%g (achieved by use of disk brakes) for operation up to 200km/h. (State of the Art 1994, p221; Technology Primer (2) 1997, p783). On vehicles fitted for Enhanced Emergency Braking, the assumed braking rate can increase to 12%g. Unfortunately, whilst the trains may be physically capable of applying sufficient brake force to achieve these braking rates, the wheel/rail adhesion required to support them is not always available.

During the 1970s, BR operated a tribometer train to record adhesion on 27 routes around the network (each varying from 1.4 to 18.6km in length). The train was fitted with two disk braked axles (the 9^{th} and 10^{th} in the direction of travel), the brakes being applied on these alternately at a gradually increasing force until the onset of wheel slip. When this occurred, the relevant forces were logged. By this means, a measurement of adhesion could be taken about every 50m at 24km/h and 200m at 97km/h. Mean adhesion rates were then determined for the route that had been covered (Pritchard 1979, pp20).

During a year of testing, mean adhesion below 15% was only recorded on 1.5% of runs. However, during wet weather it fell to between 12 and 13% in 4% of runs. Mean adhesion below 12% was only recorded on 3 occasions, when it fell to as low as 10% on normal running line and 3% for the first train running on rusted track in wet weather (Pritchard 1979, pp20-22).

During wet days in the leaf fall season, average adhesion varied from 7.5 to 18%, sometimes remaining below 10% for several hundred meters in areas with lineside trees or bushes. In the worst run, as many as three individual measurements of less than 5% were recorded (Pritchard 1979, p21).

The investigation concluded that if a 9% adhesion demand is made in wet rail conditions, 4% of the time the braking distance can be expected to increase by 8% due to areas of line with lower than 9% available adhesion (Pritchard 1979, p23).

Whilst this is not usually a problem, due to allowances made within the signalling design, there are about 400 operating incidents reported each year on Britain's mainline railways as a direct result of poor adhesion (see Table 5-1).

	1991	1992	1993	1994	1995	1996
SPADs	80	57	75	64	56	34
Platform Overruns	215	275	333	414	344	345
Collisions	1	0	0	1	0	0
Buffer Collisions	1	2	3	1	1	1

Table 5-1: Adhesion Related Incidents 1991 to 1996 (Fulford 1997, p3).

In order to reduce the occurrence of incidents such as these, a number of measures can be taken to overcome some of the limitations of adhesion.

5.4.1.2 WHEEL SLIDE PROTECTION

Wheel slide protection systems (WSP) measure the axle speed during braking in order to detect when creep is occurring. When excessive creep is detected, the WSP can release the brakes on the bogie or vehicle affected before wheel slide occurs (State of the Art 1994, p222).

First generation WSP systems use sensors and tachogenerators to measure the wheelset and train speed. If deceleration of a wheel (with respect to the vehicle) is detected, an electronic control signal is sent to an Electro-Pneumatic (EP) dump valve, which causes the axle's brakes to be released (Schofield(1) 2000, p9/26). The WSP system then begins to slowly ramp up the brake pressure again by control of the EP valve and, if the axle starts to slide again, the process is repeated until the brake is able to make the most of the available adhesion (Technology

Primer [2] 1997, p783). Unfortunately, activation of the EP valve is slow. Thus, if the control signal is sent at 60% creep, the wheel could be at 80 or 100% creep before brake release occurs (Schofield[1] 2000, p9/26). In consequence of this, the performance of first generation WSP systems is far from optimal. Indeed, they can cause such frequent application of slide control measures that braking distances are increased significantly (Kumagai et al. 1996, pp222).

The second generation of WSP systems are controlled by microprocessors that can hold creep within a very narrow band by monitoring the pseudo speed of the axle (usually by comparing 4 axles and taking the average speed) and activating an exhaust valve on the brakes when the wheel speed departs from the pseudo speed. By continuously monitoring the amount of creep that is occurring, the controller can increase the braking pressure again as soon as the creep is determined to be under control (Schofield[1] 2000, p9/26).

The rapid changes to the applied brake force that can be achieved by second generation WSP systems enables them to respond to changes in detected adhesion levels and to utilise the conditioning effects of creep discussed in section 5.4.1 in order to optimise the adhesion available to the train. Tests in the UK and Japan have shown that WSP fitted trains utilising creep in this way can achieve stopping performance better than the measured adhesion would appear to allow, with little difference in braking distance between dry and wet rail conditions (Gostling 1993, p81; Kumagai et al. 1996, pp222-3).

A third generation WSP system has also been developed in Japan. This operates on the same basic principles as a second generation system, but is designed to utilise adhesion prediction and individual-continuous control multi-position EP valves (allowing each axial brake to be controlled continuously and individually). The system is able to maintain braking at the creep saturation point (the value of creep offering peak adhesion – a region generally considered too unstable for use in second generation WSP systems) and has proved able to maintain high braking performance under snowy and icy weather conditions (Kawaguchi 1997, pp4 & 7).

In order to minimise the effects of WSP system failure, standards applied by both LUL and Railtrack have required the WSP to be arranged on a localised basis. On Railtrack, an individual component was required not to result in loss of brake force to more than one axle, whilst on LUL the WSP system must operate on a per bogie basis (Fargher[2] 2000, p5; LUL[2] 1991, p7).

Whilst WSP systems are able to maximise the use of available adhesion, they are still limited by the levels of adhesion present. "Even with optimised WSP, in worst case adhesion conditions, the adhesion may be too low for safe stopping distances to be achieved. In these situations, the available adhesion at the wheel to rail interface must be increased" (Fulford et al. 2001, p88).

5.4.1.3 VEGITATION CONTROL

Many of the most serious occurrences of low adhesion are caused by leaves on the line. The most effective method of dealing with this threat is vegetation management (Fulford et al. 2001, p85). Simply removing all vegetation from the trackside would be impracticable and environmentally unacceptable, but felling of some trees in areas where they are overcrowded or overhang the track is generally seen as good practice (Shooter 1993, pp12-18). Unfortunately, this selective felling is not enough to eradicate the problem of leaves on the line, particularly since many of the trees causing problems for railway operation are not even on the property of the railway company and, therefore, cannot be removed (Fulford et al. 2001, p86).

In areas prone to leaves on the line, alternative management techniques can be used. Chain link fences of up to 3m in height have proved effective at preventing leaves from blowing on to the line – particularly if the fence is located as close to the track as possible. Even low barriers (such as cable troughing) have been found to have a significant effect on preventing leaves from reaching the track. Another approach would be the use of low growing shrubs to act as leaf traps, encouraging leaves to accumulate away from the track (Shooter 1993, pp19-20).

5.4.1.4 ADHESION ENHANCEMENT

There are a number of approaches that can be taken to enhancing the adhesion available to a train. Some of these relate to maintenance and preparation, whilst others can be used dynamically as and when required during a train movement.

The earliest method used to enhance available adhesion was sanding, initially being introduced as an aid to traction rather than braking (Tisi[1] 2000, p42).

The first sanding systems for braking were designed to provide a one shot emergency application of sand in front of the first axle, released at a rate of 5 or 6 kg/minute for 1 or 2 minutes. Tests on rail treated with moistened paper tape showed that this type of system could increase adhesion from as little as 1% to sufficient levels for full braking. Unfortunately, the high discharge rate was also found to cause sand build up at low speeds, which could actually reduce the available adhesion due to shearing the layers of sand (Schofield[2] 2000, pp5-6). The build up of sand also caused problems with track circuits, the sand acting as an insulator between the rail and wheel (a problem that occured particularly with lightweight DMUs - the return current through the rail helping in detection of EMUs). In order to overcome this effect, the systems were developed to apply sand to the third axle of the train, thus ensuring that the leading bogie is not affected (Component Part 2002, p56).

Despite these improvements, there were still concerns about sand insulation and also that sand deposited over points and crossings could add to the maintenance required and cause additional wear and failures (Tisi[1] 2000, p44). Therefore, a general reduction in the amount of sand used was deemed desirable.

In accordance with this desire, a second generation of sanders (known as Automatic Sanding Devices) were developed to apply sand at a significantly reduced rate, of about 1.5kg/minute (Schofield[2] 2000, pp7). ASDs apply repeated automatic sanding when poor adhesion is detected by the WSP system. Sand is still applied to the third axle, both to ensure operation of track circuits and to allow continuous detection of low adhesion, with the un-sanded front bogie acting as a control for triggering the WSP activity. (Tisi[1] 2000, p42).

A third generation of 'smart' sanders are now appearing on the UK market. The 'SmartSander' system developed by AEA Technology assesses the available adhesion then delivers sand at an appropriate rate to cope with any shortfall – thus minimising sand usage when not much is required and allowing for high delivery rates when needed at high speeds (Fry 2000, p2).

The use of sand has proved highly successful. "Auto sanding under braking has been convincingly demonstrated to improve safety, relative to reducing SPADs and station over runs under low adhesion conditions as well as bringing performance benefits" (Tisi[1] 2000, p43).

In the autumn of 1999, Railtrack Southern Zone and Connex South East conducted trials on three class 465 units. All three units were equipped with low adhesion monitoring systems. One was also equipped with an AEA Technology 'Smart Sander', one with a standard ASD (used only on full service and emergency brake applications) and the third with no sanding device at all. The trial showed that the un-fitted train achieved an average deceleration rate of 3%g in worst-case adhesion conditions. The ASD fitted train achieved 4.5%g in the same conditions and the smart sander fitted train achieved 6%g (Fry 2000, p4).

If sufficient trains equipped with automatic forms of sander operate on a given line, any leaf film is broken up by repeated applications of sand in poor adhesion areas. This not only improves the adhesion available to those sander equipped trains, but also improves conditions for all other traffic operating on the line (Fry 2000, p3).

Outside of the UK, systems have also been developed to spray ceramic particles into the wheel/rail interface, instead of sand. This allows more directed application into the wheel/rail interface than can be achieved with sand, thus reducing the application rate required to improve adhesion. Laboratory tests in Japan have confirmed that adhesion can be increased by between 0.1 and 0.3 under wet conditions by use of this technique (Kumagai et al. 1996, p222).

An alternative method for applying sand to the rail is the use of sandite (sand held in a suspension of gel), which can be laid on the railhead by specially designed trains. Sandite is typically used in areas prone to low adhesion, offering an increase in adhesion of around 0.05 on damp leaf film (Shooter 1993, pp 24, 45). "Sandite has been shown to be effective in removing leaf film and will reduce the subsequent rate of new leaf film build up. However, a problematic level of contamination can still build in a matter of hours and more than one daily sandite treatment may be required" (Tunley 1999, pM13/2).

Chemical treatments can also be used to enhance adhesion. Chemicals in the form of liquids, sticks and gels have all been tested and results have shown that, in certain circumstances, these may be effective, depending on the type of adhesion control system fitted (Sergeant 1996, p88).

5.4.1.5 RAILHEAD AND WHEEL CLEANING

Rather than applying additional substances, adhesion improvements can also be achieved by cleaning contaminants from the railhead and/or wheel tread. Some of the techniques available for achieving improvements in this way are shown in Table 5-2.

In practice, combinations of cleaning techniques produce the best results. In particular, experiments in the UK have shown that high pressure water jetting is not as effective as sandite, but that water jetting to remove contaminants before sandite application is more effective than either treatment alone (Fulford et al. 2001, p86; Tunley 1999, pM13/2).

Another interesting suggestion is that as wheel creep improves adhesion for following wheels, potential improvements to braking distances could be obtained by locking up the front axle of the train and allowing it to slide along the rail, effectively scraping off contamination. It is suggested that there would be a critical speed below which this technique could be allowed to occur without causing damage to the wheel, although the author has not found any suggestion of what this speed would be (Fulford 1997, pp65-6, 88). Very little research has been done into this suggestion and the results of the testing that has been carried out appears to be contradictory. A 1981 report on tests on British Rail's Southern Region did reveal reduced stopping distances when wheel sets were permitted to lock up (Fulford 1997, p66). However, further tests in the Autumn of 1992 found that trains operating with locked wheels achieve stopping performance inferior to the measured value of adhesion (Gostling 1993, p81).

A more sophisticated approach to utilising the cleaning effect of front axles is known in Japan as 'Train Set Force Control'. This approach controls the braking effort to utilise adhesion on the axles where it is available. Higher brake forces are, therefore, applied on vehicles in the middle and rear of the train than to the leading bogies. "Field tests have demonstrated that full utilisation of the higher levels of adhesion towards the rear of the train can improve overall performance by 30%" (Kumagai et al. 1996, p221). Similar approaches have also been proposed in the UK (Broadbent 1969, pp21-2).

5.4.1.6 TRACK MAINTENANCE AND TRAIN DESIGN

In the discussion of adhesion outlined in section 5.4.1, it was noted that track irregularities (such as undulations, corrugations and curvature) can cause loss of adhesion due to reduction in wheel/rail contact force. This effect can be reduced in two ways:

- Reduction in track irregularities, by means of diligent track maintenance and/or grinding to eradicate corrugations;
- Reducing the effect of track irregularities, by using soft primary springing (between the bogie frame and wheelset) coupled with stiffer secondary springing (between the bogie and vehicle body) and, by reducing the un-sprung mass of powered wheelsets

  (Sergeant 1996, p88; State of the Art 1996, p171).

Adhesion can also be affected by wheel characteristics. Larger wheels, less weight on each wheel and softer wheel material all increase adhesion slightly by reducing the Hertzian stress of the contact zone (Sergeant 1996, p88).

Technique	Description	Results	Limitations	References
Rail Scrubbers	Rubbing a friction material along the railhead or scouring it with rotating brushes to remove crushed leaf film and moist leaves	Trials suggest improvements do occur to adhesion levels		Schofield(2) 2000, p4; Shooter 1993, pp29-31
Magnetic Track Brakes	Act as rail scrubbers, with the brake scraping along the top of the rail and cleaning off contaminants. The magnetic attraction between the pole pieces and the rail also provides a retarding effect that does not depend on the adhesion level	Can be very effective if used in sufficient quantity (usually one per rail on every bogie). However, in very low adhesion conditions the friction effect of the track brake also suffers	Limited tests of magnetic track brakes on Tyne and Wear and BRLMR suggested little benefit from magnetic track brakes in poor adhesion conditions	Fulford 1997, p35; Tunley 1999, pM13/4; Shooter 1993, p44
Controlled Wheel Creep	Slowing the rotation of the wheel with respect to the speed of the train (typically by 15 to 25%) improves adhesion and also has a cleaning effect when the wheel in 'creep' grinds against the track	The improvements in adhesion levels to be gained vary with rail conditions, but generally significant improvements are possible if controlled in a stable manner	Excessive levels of creep can cause metallurgical damage to the rail	Fulford 1997, p25; Gostling 1986, p4; Nagase 1989, p4; Schofield(1) 2000, p9/26
Auxiliary Tread Brakes	Adhesion can be enhanced by use of abrasive blocks applied to the wheel tread during braking. In the UK, this approach is generally used on disc braked trains, with up to 10% of the brake effort typically applied by means of auxiliary tread brakes in order to utilise their scrubbing and cleaning effect. In Japan, auxiliary tread brakes are designed to roughen the tread (not just clean it), even on tread brake fitted trains	In the UK, have a similar effect to controlled creep on wheel cleaning, making a small but noticeable improvement to adhesion. The Japanese approach has been found to have positive effects in moist conditions, improving braking adhesion by up to 30% on that achieved by trains with conventional cast iron brake blocks. The effects reduce if there are significant levels of contamination present	Scrubber blocks only clean the wheel and not the rail. In low adhesion conditions they, therefore, only have a small effect. Installation and maintenance has been found to be high for the short period of time each year that they are required	Kumagai et al. 1996, p221; Schofield(2) 2000, p4; Sergeant 1996, p88; Shooter 1993, p35-6; Tunley 1999, pM13/4
Plasma Torches	Designed to 'burn off' contaminants on the rail head in front of the first axle of the train.	Proved to have limited success in trials, only working well on oil	Expensive, requires a lot of power and slow transit speeds for effective results, is difficult to control and could damage the rail if the train stopped	Shooter 1993, p44; Sergeant 1996, p88; appendix A, interview with K Schofield
Laser	Also designed to 'burn off' contaminants on the rail head in front of the first axle of the train.	Does appear to act only on the contaminant in the rail head and not to affect the rail itself at all.		appendix A, interview with K Schofield
Vacuum Leaf Removal	Use of industrial vacuum cleaners to remove leaves from the area of the line	Allows clearance of large numbers of leaves	Requires operation of a special vehicle at low speeds, with the practical cleaning area limited to about 3 miles per night shift per machine	Shooter 1993, p21
Water Jets	Water is sprayed onto the rails under pressure (200litres/min of hot or cold water at 500bar in France; 80 litres/min of cold water at 1400bar in Holland)	Removes contaminants (particularly leaves and stems). In France, it is claimed that hot water works more effectively than cold. Both operators claim the systems are very effective.	Cleaning in this way makes the rail wet, so adhesion rarely increases above 0.15. It has also been claimed that high pressures jets can damage the track	Sergeant 1996, p88; Shooter 1993, p44; Tunley 1999, pM13/2

Table 5-2: Railhead and Wheel Tread Cleaning Techniques

5.4.1.7 DRIVER WARNINGS

One of the most effective ways of managing low adhesion is to "ensure drivers know of locations where it regularly occurs, the environmental conditions that can lead to low adhesion and, emergency new locations where low adhesion has arisen of which they are unlikely to otherwise be aware" (Fulford et al. 2001, p86). On an operational level, use of defensive driving techniques reduces reliance on heavy braking and thereby reduces adhesion problems (Tunley 1999, ppM13/2-3). Where drivers are aware of the need to drive defensively, including any particular areas requiring special care, the problem of poor adhesion can be largely overcome – albeit at a cost in terms of line capacity.

In an effort to assist in the dissemination of information relating to areas of poor adhesion, AEA technology have developed a low adhesion warning system designed to provide drivers with real time low adhesion warnings on a fleet wide basis. A GPS positioning system is used to identify the train's location. If significant wheel slide activity is detected on a train for more than about 50m duration, the area of poor adhesion is reported via a digital telephone network to a central base station. From there a broadcast message can be sent to all drivers advising them of the low adhesion area (Fulford et al. 2001, pp89-90; Tunley 1999, ppM13/4).

Drivers can also be advised of the need to compensate for poorly performing braking systems. In order to make use of this possibility, Railtrack standards for train braking systems stated that "where failure of a component or sub-assembly would result in the loss of more than 15% of the full service brake force throughout the train a warning shall be provided to the driver" (Fargher(2) 2000, p6).

5.4.2 ACCELERATION AND JERK RATES

In general, a passenger who finds him/herself subject to a uniform acceleration can adjust his/her body posture to compensate without much difficulty and, if the acceleration is constant, will not experience any discomfort. However, if the acceleration changes frequently, "the constant adjustments will lead to fatigue and, perhaps more important, if it changes rapidly, adjustment may be insufficient to avoid injury" (Barwell 1983, pp194-5).

Clearly, it is unacceptable for braking to exceed safe levels of acceleration or jerk (the rate of change of acceleration). It is also undesirable for normal service braking to cause discomfort to the passengers. However, in an emergency situation, it may well be acceptable and desirable for braking to reach levels that may cause some discomfort (or even minor injuries) in order to avoid an unsafe situation (such as a collision). The limits of acceleration and jerk that apply to both passenger comfort and safety are, therefore, of significant interest when determining the limits of achievable brake rates.

Unfortunately, very little research has been carried out into the physiological effects of train braking and the stresses caused by acceleration and jerk that can be safely or comfortably withstood by train passengers. The research that is available is contradictory in its findings. However, what is clear, is that the current limits applied in the specification of British rolling stock (typically up to $1.2 m/s^2$ and $1.0 m/s^3$) are not supported by scientific research (see Appendix A interview with K Schofield). Whilst limits such as these are known through experience to be safe, they are not known to represent any actual limit for either safety or comfort.

The ability of a passenger to retain balance during a brake application will be determined by a combination of factors: The magnitude of deceleration; The jerk rate; The initial posture of individual passenger (including whether seated, standing with support, standing with no support, etc.); The strength of the individual passenger; The individual passenger's speed of mental and physical reaction; How aware the passenger is of the situation (Dorn 1998, p116).

In 1932, the 'Electric Railway Presidents' Conference Committee, Bulletin No. 3' reported a test programme that exposed a large number of people to a range of deceleration magnitudes, ramps and shaped pulses. Most of the subjects were free standing (with no hand holds) and a smaller number were given handholds of various designs. On average, the subjects were found to lose their balance when:

- Free standing and facing forwards 1.61m/s^2
- Free standing and facing backwards 1.28m/s^2
- Free standing and facing sideways 1.85m/s^2
- Facing forwards and holding an overhead strap 2.25m/s^2
- Facing forwards and holding a vertical pole 2.64m/s^2

(Dorn 1998, pp117-8).

On the basis of these findings, the report recommended that a limit of 1.37m/s^2 deceleration, with a maximum jerk of 2.35m/s^3, should be applied for normal service braking and 2.35m/s^2 deceleration, with a maximum jerk of 9.0m/s^3, should be applied for emergency braking (Barwell 1983, p195)

The authors of a later study in 1972 looked at vertically moving floors, such as pedestrian conveyors. 298 adults, 156 family groups, various children and 21 disabled male subjects were studied and their response to an accelerating platform was recorded on film for later evaluation by a panel of judges. During the tests, subjects were able to hold on to a handrail. The results (shown in Figure 5-4) indicate virtually no movement at the levels typically specified in Britain. Unfortunately, the definition of 'virtually no', 'slight' 'moderate' and 'large' movement is imprecise and not defined. The author of this report remains unconvinced that levels of acceleration and jerk resulting in 'slight' or 'moderate' movement would not be acceptable to passengers, with even 'large' movement (as long as it is not so large as to cause injury) being preferable to a collision in the case of an emergency brake application. However, in the absence of a clear definition of the meaning of these terms, no definitive conclusions can be drawn from these results, shown in Figure 5-4.

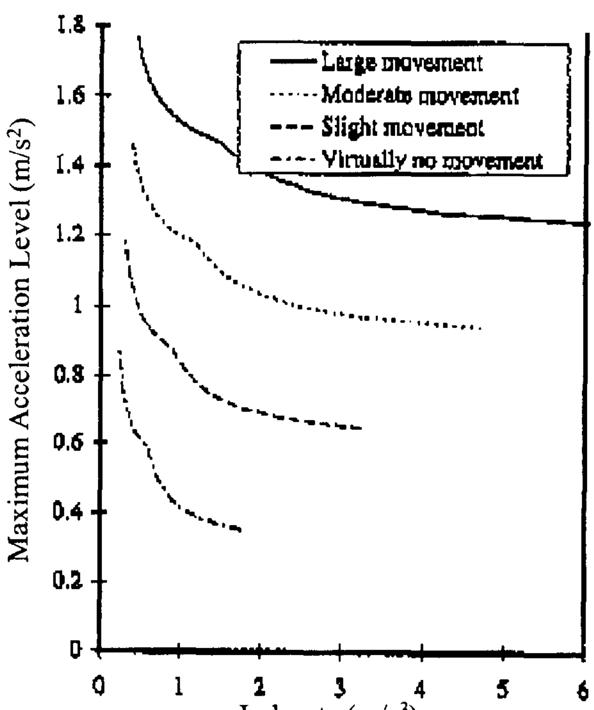


Figure 5-4: Effects of Acceleration & Jerk Rates, Browning, 1972 (Dorn 1998, pp118-9)

In 1995, a further study into retaining balance was published. This looked at 10 subjects on a short moving walkway, applying jerk values of between 1 and 10m/s^3, with a maximum acceleration of 1m/s^2. The mean values of acceleration at which free-standing subjects failed to retain posture (but not lose their balance) were found to be:

- Facing forwards 0.54m/s^2; Facing backwards 0.45m/s^2; Facing sideways 0.61m/s^2 (Dorn 1998, p119).

The instability of subjects was also found to be related to jerk, with 35% of all subjects losing posture at 1m/s^3. The study concluded that an acceptable jerk level for public transport would be below the lowest level tested and, based on previous research, a figure of 0.5-0.6m/s^3 was suggested as a realistic safe value (Dorn 1998, p119). Once again, it is not clear what is meant by 'losing posture', which appears to imply a matter of comfort rather than safety and, in the absence of a clear definition, no definitive conclusions can be drawn from these results.

Another set of tests reported in 1980 applied jerk rates of 2.45, 7.35 and 10.2m/s^3 to seated subjects. No significant effect was noted at any of these rates and the paper, therefore, concluded that jerk rates have little influence on seated passengers. This was not the case for acceleration, however. The results reported for acceleration indicated the percentage of people who 'retained' in the seat, as follows:

- Forward Facing — 50% at 3.4m/s^2, 84% at 2.94 m/s^2, 100% at 2.45 m/s^2
- Sidways facing, no armrest — 50% at 3.6 m/s^2, 84% at 2.55 m/s^2, 100% at 1.4 m/s^2
- Sidways facing, with armrest — 50% at 3.6 m/s^2, 84% at 2.4 m/s^2 , 100% at 1.4 m/s^2

  (Dorn 1998, pp119-120)

Unfortunately, the author has been unable to find the definition of 'retained in the seat' used in this research.

Based on an analysis of all of these test results, M Dorne of Frazer-Nash has suggested that the following conditions must be met in order to ensure passenger safety:

- Free standing — 0.6 m/s^2, 1m/s^3
- Standing with hand hold — 2.0 m/s^2, 3m/s^3
- Seated — 3.5 m/s^2, <2.45m/s^3

  (Dorn 1998, p125).

The figures for standing passengers with hand holds were based on the assumption that, since public transport passengers are usually unaware of the event that is about to cause braking, a reaction time of 0.75s would be required (Dorn 1998, p117).

The figures quoted by other sources also vary significantly:

- In 'Railway Engineering' it is stated that "For reasons pertaining to human physiology, maximum acceleration should not exceed 1.2m/s^2" and "Jerk should not exceed a value of 1.5m/s^3" (Profillidis 2000, p230-1);
- In 'An Introduction to Railway Braking' it is stated that, for service braking, the rate of change in acceleration is typically around 0.45m/s^3. For emergency braking the rate may reach as much as 1.79 m/s^3 before the shock to passengers (where some form of hand hold is available for those standing) becomes unacceptable (Broadbent 1969, p15);
- K Schofield, a Principal Engineer with Interfleet Technology, has stated that the maximum jerk value on UK Main Line railways is usually "0.5m/sec^3 for application and stopping at zero" (Schofield(1) 2000, p9/7). However, in discussion about this he also referred to knowledge of another reference quoting 12m/sec^3 (Appendix A interview with K Schofield);
- Research in Japan produced a series of curves for the percentage of passengers finding combinations of jerk and acceleration acceptable. These are shown in Figure 5-5.
  - The ICE high speed train operated by DB achieves an average deceleration rate of 1.4m/s^2 and peak deceleration rate of over 1.7m/s^2. Experience in Germany has also shown that passengers' freedom of movement inside a rail vehicle need not be restricted for deceleration rates of up to 2.5m/s^2 when the jerk rate is kept below 0.5m/s^3 (Braun 1979, p314-315; Braun 1993, p22);

- LUL standards require a nominal emergency brake rate of $1.3m/s^2$ on dry rails and $1.1m/s^2$ on wet rails, with instantaneous retardation limited to $1.8m/s^2$ at speeds above 20km/h and to $1.98m/s^2$ at lower speeds (LUL[1] 1991, p5). The Central Line ATO system is set to utilise service braking of up to $1.15m/s^2$ (Chandler 2001, p29);
- Light Rail Vehicles on Sunderland Direct in the UK achieve an emergency braking rate of $1.986m/s^2$ (Corser 2001, p15);
- The Sheffield Supertram achieve an emergency braking rate of at least $2.27m/s^2$ (see Appendix A, Interview with P Sharpe);
- The Manchester Metrolink type T68 light rail vehicles achieve a nominal emergency brake rate of $2.8m/s^2$, with a jerk rate of $3.1m/s^3$ (GEC Alsthom, pp45-46);
- The Railway Group Standard for managing the risks of terminal tracks requires buffer stops to retard any train type by no more than $1.47m/s^2$. For lightweight trains, this requirement is relaxed to up to $2.45m/s^2$. This implies that in an emergency (such as terminal platform overrun) the physiological effects of retardation to $2.45m/s^2$ are acceptable (Turner 1999, p5);
- The Paris Metro 'pneu' stock (which carry both seated and standing passengers) achieve up to 1.45 m/s^2 deceleration under service braking and up to $2.5m/s^2$ during emergency braking (Hardy 1988, p75);
- Bombardier have stated that the class 221 achieves a peak emergency brake rate of $1.36m/s^2$ with a brake transition time of 0.2s (Beilby 2001). That is, an average jerk rate of $6.8m/s^3$ during the brake build up.

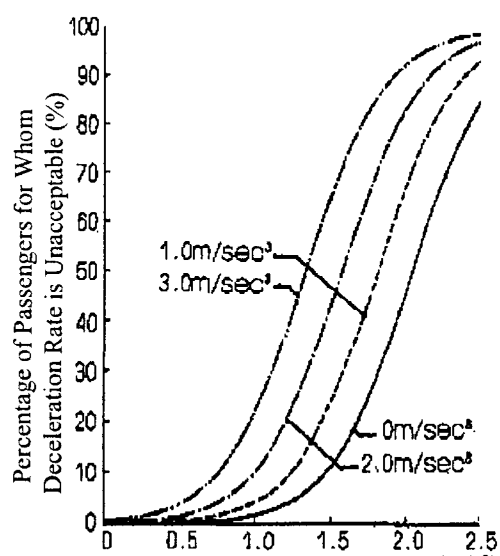


Figure 5-5: Effects of Acceleration & Jerk Rates, Study by Shiroto (Sone et al. 1998, p38)

Deceleration rates far higher than the limits of acceptability suggested by most of the research sources reported in this section have been proved to be acceptable by over 50 years of operational experience on the Paris Metro. This clearly calls into question the validity of the conclusions drawn from the research. It is unfortunate that the quantity and quality of available research is insufficient to make a definitive conclusion concerning the acceleration and jerk rates acceptable in practice for either passenger comfort or safety. However, the author believes that proven operational experience can be used to clarify the situation.

In the light of operational experience in Paris, with service brake rates of up to $1.45m/s^2$ that appear to be acceptable to the passengers, the comfort limit for acceleration must be at least as high as this. Similarly, since use of emergency braking rates as high as $2.5m/s^2$ has been proved acceptable from the viewpoint of safety, the safe limit for acceleration must be at least as high as this. As a working assumption, the author, therefore, proposes to adopt these figures as the limit of acceleration for comfort and safety.

It is unfortunate that the jerk rates applied on Paris Metro 'pneu' stock are not known by the author. However, since the limits of acceptable deceleration rates defined in the 1932 'Electric Railway Presidents' Conference Committee, Bulletin No. 3' (1.37m/s^2 for service braking and 2.35m/s^2 for emergency braking) are roughly in line with the rates used operationally in Paris, it would seem reasonable to also take the maximum jerk rates proposed in that report (2.35m/s^3 for service braking and 9.0m/s^3 for emergency braking) as a working assumption, pending the results of any more conclusive research in the future.

One point frequently raised by engineers with a main line railway background is that the safe limits of railway braking are not determined by human physiology alone. Where rail vehicles carry catering facilities, whether in the form of a trolley or a buffet car, high deceleration and jerk rates could cause drinks to be spilled, trolleys to roll away or hot fat to spill in the galley. The author has been unable to find any research considering the limits that may be imposed by considerations such as these and further research into the effects of high acceleration and jerk rates on such facilities would clearly be needed before their introduction to an operational railway. However, it is also the author's opinion that such considerations are somewhat of a red herring, based upon the current design of drink containers, trolleys and cooking facilities. As current rolling stock is not designed for increased braking performance, the design of any future rolling stock with higher braking performance would be likely to require modification. It would appear to the author entirely feasible for the re-design of rolling stock in this way to include a re-design of the catering facilities to ensure that they remain safe when used at the increased acceleration and jerk rates. After all, the potential for far higher acceleration and jerk exists on commercial aircraft that are still able to provide comprehensive catering facilities.

5.5 BRAKING MECHANISMS

A number of potential mechanisms are available for braking a train, including friction and dynamic braking that utilise the available wheel/rail adhesion, as well as other methods that avoid use of the wheel/rail interface.

5.5.1 FRICTION BRAKING

Traditionally trains were fitted with friction based braking systems, achieving retardation "by the rubbing of two components on each other to convert the kinetic energy of the train into heat" (Monselet 2001, p59). This approach is relatively straightforward and, by adjusting the force with which the two friction components are applied to each other, can be used to achieve variable braking rates. However, the use of friction in this way also has some disadvantages:

- All of the kinetic energy dissipated is converted to heat and therefore can not be reused by the train;
- Rubbing the two frictional components together causes them to wear. This results in the need for regular maintenance and associated costs for downtime, labour and materials;
- Thermal degradation of the friction surfaces and elements can occur at very high rates of energy conversion, leading in some cases to fatigue type failures.

  (Beilby 1993, p94; Watson 1999, pM07/1).

Two types of friction braking are commonly used on railways, generally referred to as tread brakes and disc brakes.

5.5.1.1 TREAD BRAKES

In tread braking, the friction required to retard the train's movement is achieved by applying a brake block directly onto the tread of the wheel (that is, to the part of a train's wheel that comes into contact with the running rail during normal motion). A representation of a tread brake is shown in Figure 5-6.

The earliest forms of railway brakes were wooden blocks applied to wheel treads. However, as speeds, train lengths and weights increased, these were unable to dissipate enough energy for effective braking. In consequence of this, cast iron brake blocks were developed in 1878. Little development in brake block materials was then made until the 1950s, when the need for enhanced brake performance led to the development of improved cast iron blocks and, subsequently, blocks made of composition materials (Watson 1999, ppM07/5-6).

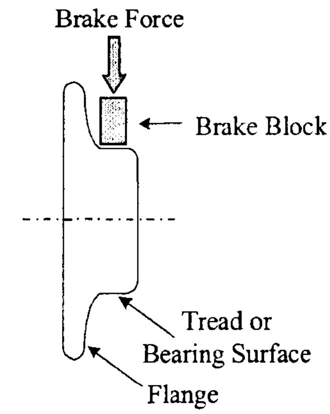


Figure 5-6: Application of Braking Force to a Tread Brake Shoe

The use of tread brakes has a number of advantages and disadvantages when compared to other braking mechanisms. These are outlined in Table 5-3.

ADVANTAGES
The brake block acts as a 'scrubber', improving adhesion by both removing built up contamination on the wheel tread and roughening its surface (Gibson 1996, p26; State of the Art 1994, pp221-2). However, 'there is no evidence to suggest that this scrubbing action will have any significant effect on heavily contaminated track as fresh contaminant is constantly being introduced into the wheel to rail interface' (Fulford et al. 2001, p87)
During braking, "wear debris can fall from the leading wheels onto the rails, further helping to improve the adhesion of the following wheels" (Gibson 1996, p26)
There is no need for a dedicated rotating part of the brake, as the wheel itself takes on this role. This has space advantages and also offers substantial cost and weight savings (Tirovic 1996, pp33-4)
Brakes acting on the wheel treads themselves are cheaper than disk brakes (Component Part 2002, p54).
DISADVANTAGES
It is only possible to apply the brake block to a limited area of the wheel surface, a 1m-diameter wheel providing not more than a 0.25 m^2 braking surface (Monselet 2001, p60)
Tread brakes can only achieve a limited energy capacity. For passenger stock travelling at 160km/h and above, the heat generated in a tread brake can affect the surface of the wheel (Technology Primer (2) 1997, p783).
The wheel rim material is selected for its rolling properties and is therefore not ideal for friction braking (Watson 1999, pM07/1)
The 'poor' wheel surface caused by tread brakes roughening the tread surface makes the vehicle noisier, particularly in the case of high speed vehicles (Tirovic 1996, pp33-4)
Tread braking can cause high wheel tread wear and accelerated wheel damage (Tirovic 1996, pp33-4)

Table 5-3: Advantages and Disadvantages of Tread Brakes

In order to overcome some of the disadvantages of tread brakes, a friction brake can instead be applied to the sides of the rim, flanges, wheel web, hub or axle. However, experience has shown that use of the flange and wheel web can lead to fatigue cracking and wheel failure. Braking on the axle involves impracticably high forces, due to the much smaller radius and can result in stability difficulties and thermal risk to the structural integrity of the wheelset. The only alternative that has been successfully developed for wide scale use in friction braking is the disk brake (Watson 1999, pM07/1).

5.5.1.2 DISK BRAKES

The disc brake, derived from automotive technology, appeared in the 1930s but did not advance until the 1950s. (Morris 1994, pRP13/7 (2)). Disks can be mounted directly on the wheel's axles, or alternatively on a drive shaft, the traction rotor shaft or the wheel (State of the Art 1994, p221; Tirovic 1996, pp34-5). "In a disc brake, two shoes, also called pads, are forced against opposite sides of a rotor, the disc. When the brakes are activated, the shoes squeeze the disc between them. The friction of the shoes and the rotor creates a retarding force, causes the deceleration of the disc. Since the disc and the axle are interdependent, the train will be retarded" (Monselet 2001, p64). A representation of a disk brake is shown in Figure 5-7.

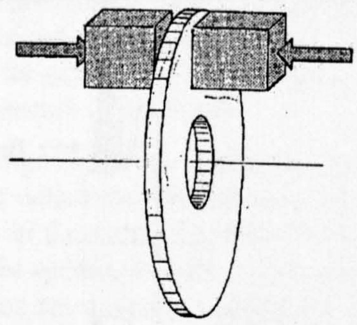


Figure 5-7 Braking Force of Brake shoes on a disc brake (Remling, 1983, p23)

The use of disk brakes also has a number of advantages and disadvantages when compared to other braking mechanisms. These are outlined in Table 5-4.

ADVANTAGES
The brake surfaces and brake pads are flat and, therefore, easy to produce (Watson 1999, pM07/1)
They can be relied upon for a good, even, contact over a substantial portion of the brake pad. This allows the rate of energy dissipation per unit area of the rotor and pad to be kept to a minimum, promoting long life (Watson 1999, pM07/1)
Their characteristics are unaffected by wear and they afford better opportunities than tread brakes for heat dissipation (Barwell 1983, p198)
Because disk brakes do not require use of the wheel tread, the design and material usage of the tread can be optimised for its primary purpose rather than braking (Watson 1999, pM07/2)
DISADVANTAGES
On disc braked trains, the wheels and rails are not cleaned by the scrubbing action previously provided by the brake blocks. This reduces available adhesion levels (Watson 1999, pM07/2; Whalley 1997, pE4/12).
Disc brakes are subject to higher temperatures than tread brakes. If care is not taken to ensure that sufficient energy can be dissipated this may lead to damage to the rotor (Monselet 2001, p65)
Disc brakes cost more than tread brakes (Schofield(1) 2000, p9/12)

Table 5-4: Advantages and Disadvantages of Disk Brakes

Fitting disk brakes to the drive shaft offers some advantages as their rotational speeds are higher than those of wheel sets, which proportionately reduces torque requirements. However, the application of drive shaft disk brakes has been limited due to problems with available space and the potential for brake failure if the transmission shaft breaks (Tirovic 1996, p34).

Wheel mounted disk brakes offer more space around the axle and simplify axle design. They can also use the wheel as a part of the heat sink, considerably enhancing power dissipation (Watson 1999, pM07/4). However, since friction can only be applied to one face of the brake, high thermal gradients can be developed, causing substantial differences in thermal expansion in the direction perpendicular to the disc friction face and resulting in disc distortion (Tirovic 1996, pp34). If heat transfer to the wheel itself occurs too rapidly, unacceptable stresses can also be placed on it, causing a gradual build up of tensile stress in the web (Watson 1999, pM07/4). As a result of these problems, wheel mounted disc brakes are usually restricted to applications with lower energy requirements (Tirovic 1996, pp35).

In some applications, disk brakes are supplemented by tread brakes that apply 10 to 20% of the braking effort. This can improve the adhesion coefficient by cleaning the wheel surface and can also prevent overheating of the disk brake system (Watson 1999, pM07/4).

5.5.2 DYNAMIC BRAKING

As technology has advanced, many train operators have adopted forms of dynamic braking, "where retardation of a vehicle is produced by a method that does not involve friction as the primary means by which the kinetic energy is dissipated" (Fargher[1] 2000, p4). Methods used have included:

- **Regenerative braking** - using the trains' motors as generators during braking, feeding resultant power back into the traction supply;
- **Rheostatic braking** - using the trains' motors as generators during braking, feeding resultant power into on-board resistor banks;
- **Hydrodynamic braking** - converting the vehicle's kinetic energy into heat in the drive shaft hydraulic fluid, for extraction by the engine and transmission cooling system;
- **Flywheel braking** - driving a flywheel mounted on board the vehicle during braking, the stored energy being re-used to drive the train's wheels during subsequent acceleration.

Dynamic brakes are good for coping with fairly high-energy dissipation for extended periods, particularly at high speeds. They also have superior controllability when compared to friction brakes. However, as they are dependent on the kinetic energy available to generate the braking force, they offer very poor or no braking effort at low speed and none at standstill. Demands for very high energy dissipation are also better dealt with by friction braking (Crawshaw 1993, p84; Monselet 2001, pp54-5; Sone et al. 1998, p36).

All forms of dynamic braking are still dependent on the wheel/rail interface. The assumed friction level between the wheel and the rail, therefore, determines the maximum dynamic braking effort that can be safely used (Beilby 1993, p97).

5.5.2.1 ELECTRIC (RHEOSTATIC AND REGENERATIVE) BRAKING

Switching the traction motors to run as generators provides a brake force due to the work done in turning the motors. This converts the train's kinetic energy into electric power (State of the Art 1994, p222). Whilst any electric traction motor is capable of producing this braking effort, their rating "clearly must be increased because of their increased duty, either by improving cooling or by making machines bigger" (Crawshaw 1993, p84).

If braking is to be achieved by converting the train's kinetic energy into electricity, the electrical power generated must be dissipated in some way. Two approaches have been developed to overcome this problem: Rheostatic and regenerative braking.

Rheostatic braking uses heating elements to achieve the required energy dissipation. The heating elements generally take the form of a bank of resistors, although some energy can also be recovered by heating the passenger compartments (Beilby 1993, p94; Component Part 2002, p54; Monselet 2001, p50). Heat can be extracted from the resistor banks by natural cooling, sometimes aided by vehicle motion or by forced ventilation (Leigh[1] 1994, p13/1(3)). The size of the resistor bank required depends on the amount of energy to be dissipated and the cooling methods available. However, where extensive use is to be made of rheostatic braking, the weight of the resistors required can become a problem (Schofield[1] 2000, p9/19).

The alternative form of electric braking, regenerative braking, returns the generated electricity to the power supply. In most applications, this is then used by other trains running on the line. However, "if there is no demand from the supply network, i.e. no train is running on the line, the line becomes 'unreceptive'. It is then impossible for the train to pass any energy back into the network" (Monselet 2001, p51). In consequence of this, it may not be possible to establish or maintain a regenerative brake and it cannot, therefore, be relied upon.

Whilst it would be theoretically possible to return the electricity to a wider supply network, such as the national grid, the author has found no references to this approach being adopted. It is instead common practice to retain rheostatic braking resistors to back up a regenerative brake (Leigh[1] 1994, p13/1(3)). In some applications friction brakes are also used as a back up in order to ensure that the rheostatic brake is never required to dissipate high levels of energy for extended periods. This allows the size of the resistor banks to be minimised. As a further option, the regenerated energy can also be dissipated in resistor banks in the supply substations when no other trains require the energy (Beilby 1993, p94).

The energy that can be recovered by use of regenerative braking can be significant. Studies have suggested that typical levels of energy recovery are "3-6% in inter city trains, 20% in mass transit and freight trains and 40% in trains on high gradients" (Profillidis 2000, p231).

The braking effort available from both rheostatic and regenerative braking is dependent on motor characteristics. The braking effort is at its maximum level when the maximum current that the motors can withstand is reached. However, at low speeds, the field current will not maintain this current level and the brake effort available is therefore lower (fading to zero effort at standstill). At higher speeds, there is also usually a limitation set by saturation of the motor field, resulting in a braking effort which decreases as the vehicle speed increases (Leigh[1] 1994, p13/1(4); Monselet 2001, p53). This is represented in Figure 5-8.

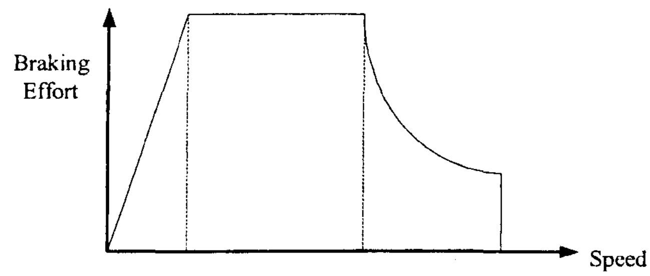


Figure 5-8: Typical Electric Brake Characteristics (based on Leigh[1] 1994, p13/1(4))

As electric braking utilises the motors, there is always a delay caused by the controlled reduction of motoring tractive effort to zero before the propulsion system can start braking. As a result "there may be a delay of up to 2 seconds before electric braking can be established" (Crawshaw 1993, p86). It is, therefore, normal to apply the friction brake transiently until the dynamic brake can take over (see section 5.6.4).

5.5.2.2 HYDRODYNAMIC BRAKING

Hydrodynamic brakes, also referred to as hydrokinetic brakes or hydraulic retarders, can be fitted to diesel hydraulic transmission systems. They consist of a rotor driven by the wheels of the vehicle and a stator that is rigidly connected to its housing. The rotor accelerates the operating fluid, which is then decelerated in the stator. In this way, kinetic energy is converted into heat in the transmission fluid during braking. The brake is independent of the diesel engine, but the heat generated is extracted by the engine and transmission cooling system. As a result, it may be necessary to run the engine above idle speed in order to ensure adequate circulation of the cooling water (Ainscough et al. 1993, p44; Leigh[1] 1994, p13/1(3); Murray 1979, p217; Whalley 1997, pE4/5).

The braking torque generated by a hydrodynamic brake increases with the square of the rotor speed. In order to achieve a constant braking torque, the amount of oil in the brake, therefore, has to be reduced as the speed increases (Ainscough et al. 1993, p44).

The hydrodynamic brake can achieve full braking effort at high speeds, limited only by the cooling system. At low speeds, however, it also begins to fade (Leigh[1] 1994, p13/1(4)).

5.5.2.3 FLYWHEEL BRAKING

Flywheel braking works on the basis of "driving a flywheel mounted on board the vehicle from the vehicle's axles via a torque converter system" (Ainscough et al. 1993, p41). This system converts the train's kinetic energy into kinetic energy within the flywheel. As the speed of the flywheel approaches equalisation with the relative speed of the vehicle, the retarding effect of the flywheel reduces and it is disconnected from the axles, continued braking being carried out by a conventional friction brake. The kinetic energy stored in the flywheel can then be reused to assist in accelerating the vehicle, thereby conserving the energy (Ainscough et al. 1993, p41).

Unfortunately, a flywheel rotating at high speeds tends to act as a gyroscope. This means that that a vehicle carrying a flywheel large enough to perform significant braking duty would tend to carry on in a straight line rather than following any bends in the track when the flywheel is rotating at its optimum speed (Ainscough et al. 1993, p41). When the flywheel is spinning at full speed (10s of thousands of rpm), containment also becomes an issue (see appendix A, interview with K Schofield).

5.5.3 WHEEL/RAIL ADHESION INDEPENDENT BRAKING

All of the braking systems discussed so far require adhesion to be available between the rail and wheel in order to achieve their potential. If insufficient adhesion is available to support an applied friction brake force, the wheel will lock. This causes the train to slide, further reducing the adhesion and resulting deceleration rate. With dynamic braking, the brake force can only be generated in proportion to the speed of rotation of the wheels. If insufficient adhesion is available to keep the wheel rotating, the brake force that can be achieved will therefore again be reduced. However, there are braking mechanisms that do not rely on the adhesion between rail and wheel. These include:

- **Mechanical track brakes** - applying a brake block directly onto the railhead;
- **Magnetic / electro-magnetic track brakes** - using large permanent or electro-magnets to apply a brake block directly onto the railhead, combined with a downward magnetic force that increases wheel / rail friction;
- **Eddy current braking** - applying a retardation force by generating eddy currents in the rail;
- **Hybrid rail brakes** - using large electro-magnets (as with track brakes) but rapidly alternating polarity, to also give an Eddy current braking effect;
- **Aerodynamic drag brakes** - increasing aerodynamic drag to decelerate very high speed trains;
- **Rubber Tyred Operation** - relying on the adhesion between a metal or concrete guide way and rubber tyres.

Brakes independent of wheel/rail adhesion are very attractive, as they help overcome the known wide variation in wheel/rail adhesion (Gostling 1986, p2).

5.5.3.1 MECHANICAL TRACK BRAKES

The first wheel/rail adhesion free braking system to be developed was the mechanical track brake. This achieves a braking effect by "the mechanical application of brake blocks, or slippers, bearing directly on the track rails" (LICS 1914, p10).

Track brakes were originally hand applied by screwing down the brake slipper on to the rail head. However, with this approach inequalities in the track surface could cause stress on the car body. In order to overcome this problem, pneumatic application was developed (with the slipper being applied to the rail by a brake cylinder when compressed air is applied and raised clear of the track by springs when the compressed air is removed). The air cylinder then acts as a cushion, preventing inequalities in the track surface from causing stress to the car body (LICS 1914, pp11-13). As an alternative to this arrangement, fail-safe application could be ensured by making the track brake spring applied and air released, but the author has not come across any reference to this arrangement being implemented.

Unfortunately, the pressure applied by the slipper brake on the rails causes the weight of the car to be lifted off of the wheels. This is a disadvantage with all mechanical track brakes, as the adhesion of the wheels is decreased when the track brakes are applied and the wheels are therefore more readily locked by the application of a wheel based brake. In order to minimise this effect, "the leverage is generally arranged so that not more than about 40 per cent of the weight can be taken on the slippers. There is then sufficient weight left on the wheels to obtain some further braking effect by means of wheel brakes" (LICS 1914, p11).

Track brakes are most widely used on low speed tram and light rail systems, but have also been used to a more limited extent on heavier rail systems. They are used, for example, at speeds between 0km/h and 30km/h on a few lines in Japan (Obara et al. 1995, p61).

5.5.3.2 MAGNETIC / ELECTRO-MAGNETIC TRACK BRAKES

Electro-magnetic track brakes utilise powerful electromagnets suspended from the train's bogies, typically at about 0.1m above the rail head. When applied, they are pneumatically lowered close to rail level and then energised to clamp the track brake to the rail head (Barwell 1983, p193; Whalley 1997, pE4/12).

The construction of electro-magnetic track brakes takes two main forms. In the original form, single pole pieces are mounted on each side of a single coil (see Figure 5-9 a). In the other form the magnet core is subdivided into a number of intermediate elements, each of which can adhere to the rail surface independently (helping to compensate for irregularities in the rail). This arrangement can be seen in Figure 5-9 b (Barwell 1983, p193). The electro-magnets are usually about 1 to 1.3m long, weighing from 160 to 200kg each (Schofield[1] 2000, p9/17).

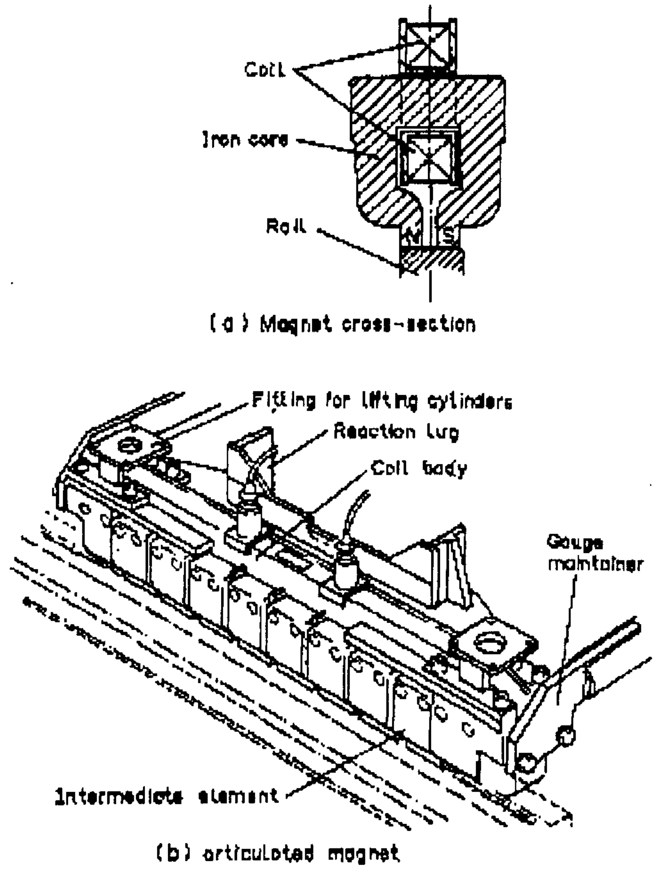


Figure 5-9: Magnetic Track Brake (Barwell 1983, p193)

A comparison of the advantages and disadvantages of electromagnetic track brakes is given in table Table 5-5.

In Germany, the electro-magnetic track brake has proved to be suitable for operation at up to 330km/h (205mph), although the braking effort reduces as speed increases. Their performance has been found to be "independent of weather conditions" and, due to the higher brake forces that they can apply when compared with brakes relying on the wheel/rail interface, they have been found to offer a higher degree of braking safety (Braun 1993, p17).

In order to reduce the costs associated with fitting magnetic track brakes to all bogies, it has been proposed that a single magnetic track brake, fitted to the leading bogie, may clean the rail sufficiently to improve adhesion for following axles. However, whilst tests on a single electro-magnetic track brake showed some effect when adhesion is above 5%, no effect was detected in 'worst case' very low adhesion conditions (Fulford et al. 2001, p87; Schofield[2] 2000, p3).

By replacing the electro-magnet with a permanent magnet, track brakes can be designed to operate without the need for an activating current. Permanent magnet track brakes are arranged so that when the brake is inactive a set of alternate North / South magnetic poles are parallel to the rail head, producing no attractive force between the brake and the rail. When applied, the magnetic poles are rotated through 90° until they are perpendicular to the rail head (appendix A, interview with K Schofield). This type of brake has a virtually unlimited source of stored energy and could, in theory, be arranged to operate in a fail safe manner (Whalley 1997, pE4/16; Appendix A interview with K Schofield). They are also easier to incorporate within bogie design (Schofield[(2)] 2000, p3).

ADVANTAGES
The magnetic force produced clamps the track brake unit to the rail head, providing a sliding braking effort and an increase in the 'effective' mass of the vehicle (Whalley 1997, pE4/12)
The sliding of the track brake unit along the rail also produces a rail cleaning effect. Tests of an electromagnetic track brake fitted train in Japan confirmed "that the frictional coefficient on rails after passing was 10% higher than before passing" (Obara et al. 1995, p63; Whalley 1997, pE4/12)
The mechanical braking effect can be varied by adjusting the current applied to the magnet's exciting coils and thus adjusting the attractive force with which the track brake is drawn down upon the rails (LICS 1914, pp18-19)
Requires much lower current than an eddy current brake (see section 5.5.3.3) (Obara et al. 1995, p63)
DISADVANTAGES
The weight of a track brake makes a significant addition to the usual bogie mass (which is arround 5 to 8 tonnes) which, together with the forces that they generate during braking, leads to the need for a redesign of the bogie to cope. This makes them expensive to install and maintain (Schofield[(1)] 2000, p9/17; Schofield[(2)] 2000, p3; appendix A, interview with K Schofield)
Only a small contact area with the rail can be achieved, an air gap parallel to the rail being required between the N and S poles. This makes stabilising the brake force of the magnetic track brake difficult (Obara et al. 1995, p63)
The rubbing element is generally made of cast iron and has a rising friction characteristic with decreasing speed. When used on high speed trains, design to ensure a high brake rate at high speeds prevents their normal use at low speed - or the heat generated would be excessive. Similarly, if designed for low speed application their braking effort will be insufficient at high speeds (Barwell 1983, p193; Obara et al. 1995, p62; Schofield[(2)] 2000, p3)
Magnetic track brakes do not work on high manganese rail (Gostling 1986, p2)
They may interfere with signalling as they deposit debris and induce high magnetic fields. They can also damage track and junctions (Schofield[(1)] 2000, p9/17)
Steel track brakes are the most effective, but when used at speeds above 150km/h they tend to pick up steel from the track. This then has to be cleaned off or it holds the brake away from the rail and reduces its effectiveness (Berndt et al. 1979, 230). Whatever material is used for their manufacture, the contact surfaces of the magnetic track brake require maintenance attention to correct for wear damage (Braun 1993, p17).

Table 5-5: Advantages and Disadvantages of Electro-Magnetic Track Brakes

5.5.3.3 EDDY CURRENT BRAKING

The eddy current brake differs from the track brake in that there is no actual contact between the rail and brake shoe. Aan air-gap of about 7mm is usually maintained, the achieved braking force reducing by about 10% for every 1mm that the air gap is increased. The magnetic segments are wound with alternate north and south polarities, such that flux passes across the air-gap, along the rail and then across to the other pole (see Figure 5-10). When there is no relative motion between the brake and rail, there is simply a normal force between them. However, when there is a relative motion, voltages are generated, resulting in the formation of eddy currents. These distort the magnetic field and introduce a transverse component in the force between them. The brake is arranged such that these forces oppose the train's motion, providing a brake effort that converts the kinetic energy of the train into thermal energy in the rail (Barwell 1983, p194; Saumweber et al. 1986, p3; Schofield 2000[(1)], p9/17).

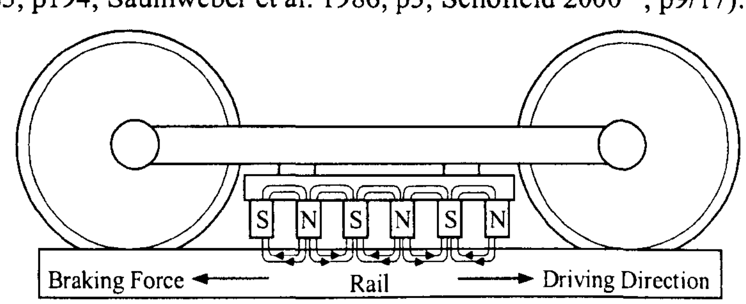


Figure 5-10: Eddy Current Brake

Tests have revealed that a 1.08m eddy current brake produces its peak retardation rate at about 75km/h (47mph). This falls off sharply as speeds decrease and more gradually as they increase (Barwell 1983, p194). A comparison of the advantages and disadvantages of eddy current brakes is given in Table 5-6.

It is possible to envisage a sophisticated train control system capable of imposing speed restrictions related to rail temperature, in order to avoid the problem of rail buckling (Gostling 1986, p3). However, the advantage of any such system has yet to be demonstrated (and it would be liable to introduce inconsistency to train control performance).

ADVANTAGES
The braking effect produced is independent of rail adhesion.
They make no contact with the rail and do not wear, making them suitable for a service brake application (Berndt et al. 1979, p231)
Magnetic attraction increases the apparent axle load whilst the brake is applied, therefore increasing the level of adhesion braking that can be achieved (Obara et al. 1995, p62)
DISADVANTAGES
As energy is dissipated as heat in the rails, there is a risk of overheating and buckling the rails (Barwell 1983, p194; Obara et al. 1995, p62; State of the Art 1994, p223). In service, they increase the rail bulk temperature by 20-30°C in specific braking sites and the surface temperatures momentarily rise by 70°C. They, therefore, require sophisticated control to ensure that the track is not damaged due to localised heating (Schofield 2000[(1)], p9/17)
They form a significant addition to the usual bogie mass, having a similar weight to electromagnetic brakes (Schofield 2000[(1)], p9/17)
They generate circulating currents in the rails, creating the potential for interference with track circuits and other signalling equipment (Whalley 1997, ppE4/15-16). Test results on both DB and SNCF confirmed that no interference was experienced when travelling over switches or track circuits. However, effects were found with axle counters and magnetic contacts (although subsequent bench tests revealed that these effects could be countered by fitting magnetic guide plates to the pole coils) (Saumweber et al. 1986, p3)
Require high current (much higher than an electromagnetic track brake) to generate braking force (Gostling 1986, p2; Kumagai et al. 1996, p223; Obara et al. 1995, pp62-3). This makes it difficult to construct the eddy current brake systems with only batteries supplying the current and, therefore, generally precludes their use as an emergency brake – although they have been used for that purpose in Germany (Obara et al. 1995, p62; Saumweber et al. 1986, p2).

Table 5-6: Advantages and Disadvantages of Eddy Current Brakes

As with the magnetic track brake, a permanent magnet version of the eddy current brakes has been developed to overcome the limitations associated with power demand (Berndt et al.1979, p231).

5.5.3.4 HYBRID RAIL BRAKING

Comparison of the properties of magnetic track brakes and eddy current brakes reveals advantages and disadvantages that could be largely overcome by using them in combination. In accordance with this view, a hybrid rail brake was developed in Japan to incorporate the features of both eddy current and magnetic track brakes. This consisted of a rail brake in which the magnetic poles of the track brake alternate between N and S in the driving direction as in an eddy current brake, but are still brought into contact with the rail as in a magnetic track brake. Both frictional and magnetic forces can then be utilised in the same brake, whilst the absence of an air gap means that much lower power levels can be used than would be required for a conventional eddy current brake (Obara et al. 1995, p61).

The hybrid rail brake, also referred to as an 'absorption eddy current rail brake', requires one tenth of the current of an eddy current brake to obtain an equal braking power and has been found to have little effect on signalling equipment (Kumagai et al. 1996, p223). Tests have also shown that the heating effects in the rails remain below the buckling strength of the rail and thus represent no danger of buckling (Obara et al. 1995, p64).

In August 1992, tests were conducted on a train fitted with both a hybrid rail brake and conventional adhesion brake in order to assess possible brake shoe materials. During emergency braking, a material offering high permeability and low friction coefficient was found to enable sufficient braking force to achieve deceleration rates of over $1.6 m/s^2$, whilst a material with low permeability and a high friction coefficient enabled deceleration rates of over $1.7 m/s^2$. With both materials, these rates could be achieved in both dry and wet rail conditions and did not depend much on speed (Kumagai et al. 1996, p224; Obara et al. 1995, pp63-4).

5.5.3.5 AERODYNAMIC DRAG BRAKES

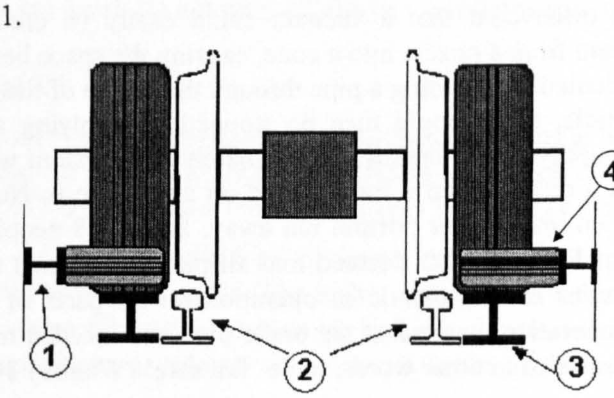
The author has only found one reference to the use of aerodynamic drag braking but, due to the potential that appears to be offered by this approach, considers that the inclusion of a description is warranted.

All train movements are affected by drag effects. When a train is accelerating these effects are undesirable and must be overcome by the propulsion system. In the interest of efficiency, it is, therefore, desirable to reduce the effects of drag. However, in braking, drag assists in the retardation (particularly at very high speeds). An example of this can be seen in Figure 5-14 in section 5.7. Aerodynamic drag braking would operate on the basis of increasing the aerodynamic drag of a very high speed train, in order to cause deceleration when required, much in the same way that aircraft utilise adjusted flap positions on landing. "This process will be limited by noise and turbulence considerations, but for a speed of 300km/h, a deceleration of 4%g appears possible" (Gostling 1986, p3).

5.5.3.6 RUBBER TYRED OPERATION

The first rubber tyred metro vehicle was built for the Paris Metro in 1951. Since then, lines 1, 4, 6, 11 and 14 of the Paris Metro system, the Lille and Marseille metros (France), Tokyo, Hiroshima and Sapporo (Japan), Santiago metro (Chile) and all lines of the Mexico City metro (Mexico) have also been developed for rubber tyred ('pneu') operation (Garbutt 1997, pp9, 82, 111, 114 ,119; Hardy 1988, p79).

The track on the Paris lines consists of conventional steel rails that act as 'safety' rails. These are flanked by longitudinal bearing strips of metal or concrete, on the outside of which vertical guide bars are installed (which act as both guides for train movement and conductors for the power supply). The trains bogies are fitted with rubber tyred wheels that run on the bearing strips and carry the weight of the train. On the inside of these tyres, conventional steel wheels are fitted. These have deep flanges and are designed to drop onto the steel rails if the tyre deflates. Horizontal rubber guide wheels are also mounted on the bogies to keep the train within the guide bars (Garbutt 1997, p45; Hardy 1988, p74). A representation of this arrangement can be seen in Figure 5-11.



1: Guide track; **2** : Rails; **3** : Bearing strip
4: Horizontal guide wheel

Figure 5-11: Arrangements for Rubber Tyred Operation (based on Peigne-Stage 2002, p2)

At points and crossings the vertical guide bars are interrupted and the height of the bearing strips is lowered, so that the steel wheel drops onto the rail. The train then traverses the crossing as a conventional train, before resuming pneumatic tyre running at the other side (Hardy 1988, p74).

By the use of this arrangement, much better adhesion levels can be achieved than with a steel wheel running on a steel rail. As a result, much higher acceleration and brake rates can also be achieved. In Paris, these are $1.3m/s^2$ for acceleration, $1.45m/s^2$ for service braking and up to $2.5m/^2$ for emergency braking. These rates are not significantly affected by wet weather, but problems can be experienced when snow settles on the bearing strips (Garbutt 1997, p15; Hardy 1988, p75).

The main disadvantage of rubber tyred systems is the initial costs, with a total of six beams and rails required in a single track and each bogie having 12 rubber / metal wheels. Operation on rubber tyres also substantially increases power consumption and leads to high level of heat generation, which can cause ventilation problems (Garbutt 1997, p15).

On the Paris metro, the tyres are changed about every 450,000km, with visual checks being made at each maintenance session. The rate of flat tyres on all 4 lines is 1 a month at worst (Hardy 1988, p79).

The Sapporo system, opened in 1971, is a simplified version of the Paris 'pneu' system. It uses only one central guide-rail rather than side-rails, reducing the amount of track infrastructure required. Interestingly, as Sapporo is prone to a lot of snow, the whole line is encased in a snow shield to avoid adhesion problems (Garbutt 1997, pp16, 46, 119).

5.6 BRAKE ACTUATION

The two most important factors in controlling a train of vehicles are:

(a) Safety - if the train should break in two, the brakes should automatically apply on both parts;

(b) Controllability - the driver's commands to apply and release should be responded to as fast as possible but the brake forces generated at the front and rear should not differ by too great a margin"

(Cable 1994, p12/2(2)).

The actuation of early braking systems was fairly simple, relying upon the vigilance of the engineman and the brakesman (Whalley 1997, pE4/1). "A brake shoe could be pivoted on an arm over the wheel treads. A horizontal rod connected the free ends of the pivot arms, so that when the rod was pulled the brake blocks were applied to the wheels. To apply the brake, a wheel was turned in the cab or the tender. By means of a screw thread, rotary motion was converted into the pull on the rods" (Technology Primer (2) 1997, p780).

It was then discovered that a vacuum could easily be created on a steam locomotive by directing steam from a nozzle into a cone, causing the space between the nozzle and the cone to become evacuated. By running a pipe through the length of the train, coupled to brake cylinders on each vehicle, trains could then be stopped by applying steam to the ejector, creating a vacuum and applying the brakes. Unfortunately, this system was not originally designed to be fail safe. As a result, when a train stalled on an incline in Northern Ireland in 1889 and was divided, the un-braked rear portion ran away, killing 78 people. In response to this, the 1889 Regulation of Railways Act decreed that all passenger trains must be fitted with a brake that was "continuous and automatic in operation on all parts of the train. The requirement for 'automatic' operation meant that the brake pipe now needed to be evacuated or pressurised to release the brake, or in other words, to be 'fail safe' (Whalley 1997, pE4/1).

5.6.1 VACUUM AND AIR ACTUATION

In 1871, George Westinghouse developed the 'air brake', providing the brake force by means of compressed air acting on a piston in a cylinder. In an air brake system, a pipe can be run along the train, connected between vehicles by use of hoses in order to allow for some relative movement between them (Technology Primer (2) 1997, p780). A compressor on the locomotive then charges a main reservoir to a pressure of 7 to 10 times atmospheric pressure (7-10 bar), which in turn charges the brake pipe to around 5 bar. On each vehicle in the train a spur connects the brake pipe to a piston valve, which is also connected to the brake cylinders and an auxiliary reservoir. When full brake pipe pressure is applied, the piston valves allow air from the brake pipe to fill the auxiliary reservoirs. However, when the driver operates the brake, some air is able to escape from the brake pipe, dropping its pressure. This drop in pressure causes the piston valves to move, cutting the auxiliary reservoirs off from the main brake pipe and connecting them to the brake cylinders instead. Thus the brakes are applied and maintained by the pressure of the auxiliary reservoir (Cable 1994, pp13/2(3-4)).

When the driver releases the brake, the brake pipe pressure builds up again. As soon as the pressure rises above that of the auxiliary reservoirs, the piston valves return to their original position, allowing the air in the brake cylinders to exhaust. This fully releases the brake and also allows the auxiliary reservoirs to recharge to their working pressure (Technology Primer [(2)] 1997, p780).

An alternative system developed shortly after the Westinghouse air brake is the vacuum brake. This again utilises a brake pipe running through the train, connected to an exhauster on the locomotive that sucks air from the brake pipe, producing a partial vacuum. Cylinders are then connected to the brake pipe, such that the vacuum causes the cylinders to release the brakes. A non-return valve in a piston within each cylinder then opens, allowing a partial vacuum also to develop on the other side of the piston. To apply the brakes, the driver simply lets air into the brake pipe. Because the non-return valve retains the vacuum in the cylinder, the difference in pressure created at each cylinder causes the piston to operate, applying the brakes. The brakes can then be released again by reconnecting the brake pipe to the exhauster (Technology Primer [(2)] 1997, pp780-1).

As the vacuum brake operates at lower pressure differentials than the air brake, it needs larger components to achieve the same brake force. Theoretically, a 21-inch diameter vacuum cylinder can produce a pull of 22.4kN (Technology Primer [(2)] 1997, p781). In practice, however, three factors further reduce the maximum force available by use of vacuum brake actuation:

(a) The vacuum exhausters cannot extract all of the air from the cylinder. In practice, this means that the brake pipe pressure is only reduced to around 0.3 bar at the front of the train and, due to leakage, around 0.42 bar at the rear of the train;

(b) As the piston rises in the cylinder, the volume of the chamber is reduced and the pressure in the chamber rises by approximately 0.1 bar, reducing the pressure difference across the piston;

(c) If the cylinder is to operate at high altitudes, atmospheric pressure is reduced below the 1 bar available at sea level. At a height of 1000m normal pressure is 0.9 bar and at 2000m it is 0.8 bar;

(Cable 1994, pp12/2(1-2)).

These factors can reduce the force available on a vacuum braked train by as much as 60 to70% of the theoretical value (to around 15.7kN). In contrast, a 16 inch diameter pneumatic brake cylinder at a pressure of 3.5 bar gives a force of 45.5kN (Technology Primer [(2)] 1997, p781).

Increased expense is not the only effect of the larger components required due to the lower pressure differentials in a vacuum brake. The higher volume of air that needs to be moved to achieve a given brake force also means that an air brake application or release can be achieved more quickly than application or release of a vacuum brake of the same pipe length (Cable 1994, p12/2(5)).

Despite these drawbacks, the vacuum brake was selected as the standard system for use on UK railways in 1881 (with use of the Westinghouse brake continuing in the USA and Europe). The explanation for this perhaps lies in the fact that, in the days of steam engines, a vacuum could be obtained at little cost by use of a steam ejector. In contrast, the Westinghouse brake required a compressor, main and auxiliary reservoirs. The vacuum cylinder was also simpler than the Westinghouse control valve and did not require a separate brake cylinder (Technology Primer [(2)] 1997, p781).

Whilst both the air and vacuum brake are theoretically fail safe, since the brake applies if the brake pipe is broken, the air brake requires that the auxiliary air reservoir (sometimes referred to as the direct air supply) is charged and it will not apply if it is expended. In consequence of this, "loss of the vacuum will automatically apply the train brake. Loss of a direct air supply will not automatically apply the brake" (British Rail, undated, p19). It may, therefore, also be that the UK decision in favour of vacuum brakes reflected concerns over the failure modes of the air brake. However, when it is considered that both the vacuum and air brake require some source of stored energy (in the auxiliary reservoir for an air brake and in the cylinder itself for a vacuum brake) that will probably have leaked away "On a new vehicle, or one that has been standing for some time" , this seems unlikely (Whalley 1997, pE4/2).

As traction systems developed and the use of steam (required by the exhauster) became less common, the vacuum brake lost most of its cost advantages. Indeed, as most modern trains began to use compressed air to supply the suspension and door operation systems, the provision of air actuated brake systems became more economical (State of the Art 1994, p222). Air braking has, therefore, tended to supplant vacuum braking.

The safety and controllability requirements of a train brake can be met by fitting a vacuum or air brake cylinder to each vehicle in the train (Cable 1994, pp12/2(2-3)). However, propagation delays in both systems mean that simultaneous application of brakes on all vehicles is not possible (Profillidis 2000, p231). "An automatic air brake theoretically propagates itself down the brake pipe at the speed of sound (1077ft/sec). The delay between the front and rear of a 13 coach train should, therefore, only be 1 second. In practice, because of the changes of direction that the air makes when negotiating bends in the pipe work and valves, ...the propagation time can extend considerably and delays exceeding 10 seconds are not unknown" (Whalley 1997, pE4/4).

Faster application times can be achieved with air brakes by fitting valves on each car that respond to an initial drop in brake pipe pressure by allowing a small gulp of air to be removed locally, speeding up the signal propagation. Even more rapid response can be achieved by installing larger valves to connect the brake pipe to atmosphere when a rapid change in pressure is detected. However, this makes the application rate uncontrollable (Cable 1994, p12/2(5)).

In a similar way, the speed of application of a vacuum brake can be increased by fitting a 'direct admission' valve to each vehicle. This valve senses that pressure in the train pipe is increasing and allows air from the atmosphere into the train pipe locally (Monselet 2001, p46).

If a faster and more controllable response is required, it becomes necessary to distribute brake release valves along the train and remotely operate them by electric signals (State of the Art 1994, p222). This approach is referred to as Electro-Pneumatic (EP) actuation.

5.6.2 ELECTRO-PNEUMATIC AND ELECTRIC ACTUATION

"With a perfect brake the retarding force should start simultaneously on all axles and rise to its controlled value at the same rate. With an electro-pneumatic or electric brake this perfection is possible of attainment." (Broadbent 1969, p21).

The Electro-Pneumatic (EP) valve incorporates a solenoid that opens and closes the valve when fed by electricity. By locating EP valves in each vehicle, or each bogie, propagation delays along the brake pipe can be minimised and brakes can be applied and released effectively instantaneously (Technology Primer (2) 1997, p781). Control signals can be sent as digital signals (either directly or in a coded form) or as analogue signals (proportional to a DC voltage, current or pulse width modulation). An analogue result is also sometimes obtained by using digital signals which are operated in a time dependent manner (Leigh(1) 1994, p13/1(1)).

The control signals can be arranged in an 'energise to apply' or 'energise to release' form, depending on the railway and the service being undertaken by the vehicles. 'Energise to apply' is not fail safe as regards the basic signal and is generally used in conjunction with a monitoring system which will change the brake system over to a back up/shadowing automatic air brake or a fail safe emergency brake. "Energise to release is fail-safe as loss of signal applies the brakes, however, it is necessary to provide some way to release the brakes in a failure situation to enable the train to be moved" (Leigh[1] 1994, p13/1(1)).

In the 1980s, "British Rail Research developed an electric brake-actuator, where electric motors driving screw jacks applied force to the brake pads" (State of the Art 1994, p222). Systems operating on this basis have yet to be proved, and it is not clear how they could ensure application in the event of power failure, but they could be of interest in the future.

5.6.3 HYDRAULIC ACTUATION

Hydraulic & electro-hydraulic systems operate at much higher pressures than air or vacuum brakes (up to 100bar), which requires greater care and cleanliness in maintenance. The cylinders are light weight, compact, about a third of the mass and much more reliable than a conventional pneumatic system. The response time of hydraulic brakes is also faster than air brakes (~0.2 seconds), giving greater control of the wheelset (although this response is sometimes too quick and has to be limited) (Kumagai 1996, pp224-5; Schofield[1] 2000, p9/24).

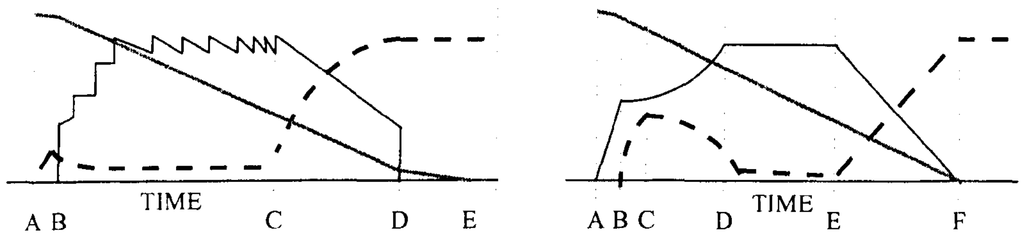
Whilst emergency braking has been provided by a completely hydraulically applied system using hydraulic accumulators, there is concern that leakage of the hydraulic fluid could cause brake failure. The initial cost of hydraulic systems is higher than air based systems. However, the overall cost of the train may prove to be lower due to the smaller component size (Schofield[1] 2000, p9/24).

5.6.4 BRAKE BLENDING

Due to delays in reconfiguring control circuits from traction to braking, the initial application time of electric dynamic braking systems tends to be longer than that of a friction brake. Similarly, initial application of a hydrodynamic brake suffers from a delay whilst the retarder fills with sufficient oil for any braking effort to be achieved. The braking effort available from dynamic braking systems also reduces as the train's speed decreases and can disappear completely. However, once established, their response to changes in required braking rate is much faster than that achievable with air operated friction brakes (although similar to hydraulic friction brakes) (Leigh[2] 1994, p45). It is desirable to utilise dynamic braking in order to minimise brake wear and make use of the controllability offered by their rapid response to changes, yet delays in application cannot always be tolerated. An optimal brake application would, therefore, consist of a blend of the characteristics of both friction and dynamic braking. Electrically controlling brake application, as well as providing the ability to remotely control friction brake systems by use of EP valves or electric brake actuators, enables the processing of brake demands and 'blending' of friction and dynamic brakes to give the desired total braking effort throughout braking.

There are two main forms of brake blending, 'Restricted Application' and 'Continuous Blend'. Restricted application blending calls for both the friction and dynamic brake at once, reducing the friction brake application to a restricted level once the dynamic brake has established itself. This condition is maintained until the dynamic brake falls to a predetermined level, when it then fades away at a controlled rate. The friction brake then applies in its place. This is the oldest style of brake blending, and is usually applied without any electronic control. Continuous blending gives precedence to the dynamic brake, using electronic control to vary the friction brake application continuously against variations in the dynamic brake (Monselet 2001, p56). A comparison of their application is shown in Figure 5-12.

It is not possible to guarantee that dynamic brake equipment alone will be able to brake a train to, and hold it at, standstill under all fault conditions. However, if the dynamic brake were to fail at low speed with the friction brake off, the delay in obtaining a friction based emergency brake could be sufficient to cause an overshoot. It is, therefore, normal practice for blending to fade out the dynamic brake below 10km/h (8km/h on LUL), so that the friction brake is fully on by the time that the train reaches standstill (Crawshaw 1993, pp86-7; LUL[1] 1991, p7).



Restricted Application Blending

- A Brake application initiates
- B Dynamic brake establishes, friction brake reduces to fixed holding level
- C Dynamic brake reaches last notch and begins to fade at known rate
- D Dynamic brake fully released. Friction brake fully applied
- E Train at standstill

KEY
- Dynamic Brake Retardation Force
- Friction Brake Retardation Force
- Train Speed

Continuous Blending

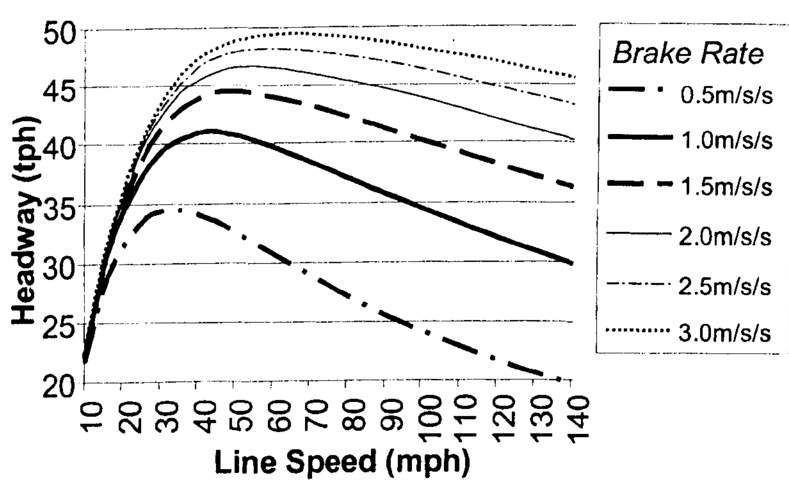
- A Brake application initiates
- B Dynamic brake builds up to maximum permissible at particular speed. Friction brake takes up braking deficiencies
- C-D Speed reduces, dynamic capability increases, friction brake decreases proportionately
- E Minimum speed for full dynamic brake reached. Dynamic starts to fall, friction brake increases proportionately
- F Train at standstill. Dynamic brake zero, friction brake fully on

Figure 5-12: Comparison of Restricted & Continuous Brake Blending (Leigh[(2)] 1994, pp43-4)

The idea of blending different types of brake system has been extended in some cases to include 'cross blending'. That is, where the dynamic braking effort available on motored axles is in excess of that needed to brake the motored axles alone, the approach utilises surplus dynamic braking effort to brake the un-powered axles as well. This technique has the advantage of minimising brake wear on the un-powered axles, but can cause problems if insufficient adhesion is available to the powered axles. The use of this technique therefore needs to be limited, and accompanied by WSP systems that can detect slide and reduce the amount of cross blending if it occurs (Leigh[(1)] 1994, p13/1(6); Whalley 1997, ppE4/9).

5.7 BRAKING PERFORMANCE

The braking performance achieved by a railway's trains has a major impact on the capacity that can be achieved by that railway. An example of this can be seen in Figure 5-13.



Assumes Moving Block Operation with: 80m safety margin; Train Length 200m; Sighting Distance 275m; Dwell 30s; Acceleration 1m/s/s

Figure 5-13: Headway Against Line Speed for Different Brake Rates

British main line multiple units fitted with a 3 step service brake would typically be designed to achieve the braking performance outlined in Table 5-7.

Step	Deceleration	
1	$0.29 m/s^2$	3%g
2	$0.58 m/s^2$	6%g
3 (full service)	$0.88 m/s^2$	9%g
4 (emergency)	$1.17 m/s^2$	12%g

Table 5-7: Designed Deceleration rates for UK Main Line Multiple Units (Tisi[(2)] 2000, p4)

Once built, the trains are tested against these deceleration rates under dry rail conditions only. There is no performance specified for low adhesion conditions. Clearly, this means that the braking rates actually achieved in poor adhesion conditions could be far lower. Whilst the safety of trains specified to these braking performances has proved adequate historically, they have been operating on lines fitted with conventional signalling systems that usually allow a 25% contingency between the signal spacing and trains' theoretical stopping performance. This contingency is sufficient to a allow for a small amount of over speeding by drivers and for poor adhesion performance (see appendix A, interview with N Harwood). With the introduction of more advanced train control systems operating in moving block, protection against drivers over speeding will be achieved through ATP based speed supervision, potentially reducing some of the margin required. However, allowance will still need to be made for poor adhesion within the calculations of safe train separation distances. This could be achieved by repeating the conventional signalling approach of adding a significant margin to theoretical braking distances, or could be managed by developing a better understanding of worst case achievable braking rates and supervising against them.

It has been recognised for some time that high density urban rail systems operating at speeds up to 96km/h (60mph), utilising composite block tread brakes only, can achieve emergency brake deceleration rates of up to $1.43 m/s^2$ (Broadbent 1969, p13). On LUL, the specification of emergency braking performance is not this high, but does extend beyond dry rail conditions. A nominal brake rate of $1.3 m/s^2$ in dry rail conditions and $1.1 m/s^2$ in wet rail (continuous heavy rainfall) conditions is specified (LUL[(1)], p5). Whilst this does not account for worst case conditions (at the start of rainfall, during frost or with oil based contaminants on the line), it clearly provides an increased confidence that safe braking can be achieved under sub-optimal conditions. However, the question still remains as to whether better braking rates could be practically achieved.

In order to improve the braking performance of trains, "there are two aspects to be addressed: adhesion and braking force" (Kumagai et al. 1996, p219). Both of these aspects have been addressed on Sunderland Direct, by utilising electromagnetic track brakes with a traction battery back-up in addition to wheel (disk or tread) brakes. As a result, the quoted emergency braking rate for their vehicles is $1.986 m/s^2$ (Corser 2001, p15). On Manchester metrolink, disk, track and electric brakes have all been combined with sanding to achieve an emergency braking rate from 50km/h or less of $2.8 m/s^2$ (GEC Alsthom, p45). This approach has commonly been adopted on Light Rail Vehicles. Typically they can be expected to achieve an emergency brake rate of $1.3 m/s^2$ utilising the wheel/rail interface alone, but this "can be easily enhanced to 3.0m/s/s" by use of electro-magnetic track brakes (Whalley 1997, pE4/13).

Elsewhere in Europe, higher brake rates have also been achieved in operation on main line and metro vehicles by addressing both the aspects of adhesion and braking force. By blending different brake mechanisms (including magnetic track brakes), the German high speed ICE trains achieve a deceleration rate in emergency braking that averages over $1.4 m/s^2$ and exceeds $1.7 m/s^2$ at some speeds (see Figure 5-14). By addressing adhesion in a different way, the rubber tyred 'pneu' stock on Paris Metro achieve $1.45 m/s^2$ deceleration under service braking and up to $2.5 m/s^2$ during emergency braking. Rates that are achieved entirely through the rubber wheel / guide way interface and are not significantly effected by wet weather (Hardy 1988, p75).

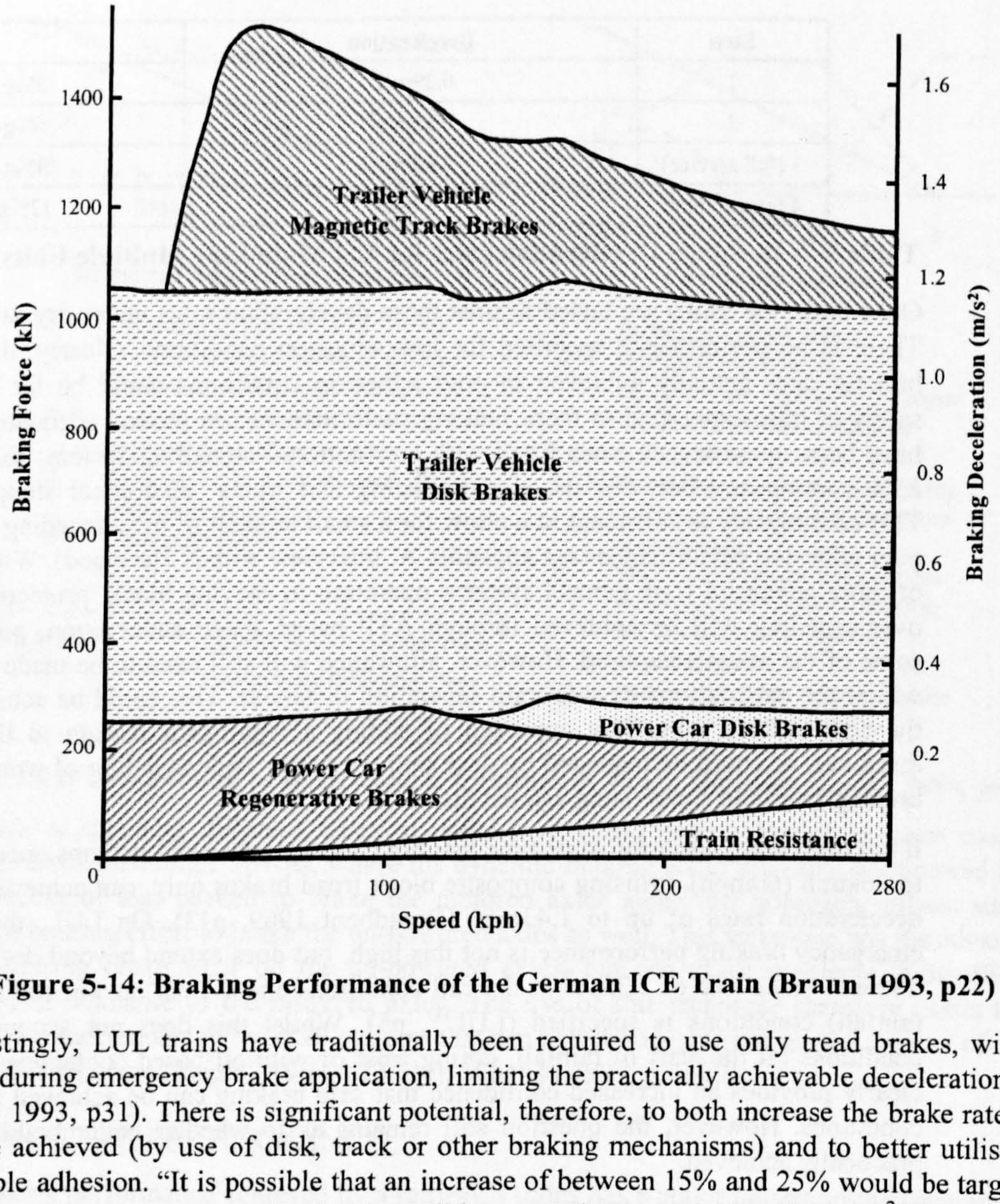


Figure 5-14: Braking Performance of the German ICE Train (Braun 1993, p22)

Interestingly, LUL trains have traditionally been required to use only tread brakes, without WSP, during emergency brake application, limiting the practically achievable deceleration rate (Rowe 1993, p31). There is significant potential, therefore, to both increase the brake rate that can be achieved (by use of disk, track or other braking mechanisms) and to better utilise the available adhesion. "It is possible that an increase of between 15% and 25% would be targeted, raising retardation rate – under 'dry' conditions – to between 1.55 and 1.7m/s^2". In order to ensure safe operation at brake rates such as these in poor adhesion conditions, "would require the provision of wheel slide protection equipment"; an approach that could be safely implemented by designing the WSP system to operate on a per axle basis (Rowe 1993 p31).

This approach has been adopted on the express railway lines in Japan to allow an increase in operating speeds from 110 to 140km/h without increasing braking distances. A third generation WSP system utilising adhesion prediction and continuous individual control of each axial brake has been fitted to tread braked trains. The brake shoe material has also been changed to an alloyed cast iron, offering very high (but unstable) braking characteristics capable of producing an emergency brake rate of up to 2.9m/s^2. The advanced WSP has been used to control the unstable characteristics of the brake shoes and at the same time reduce passenger discomfort, by monitoring and rapid control of the train's braking performance. As a result, the two classes of train fitted with this system (the class 281 and 283 DMUs) have been shown to reliably achieve a nominal brake rate of 1.5m/s^2 when braking from 140km/h (87mph) under snowy and icy weather conditions (Kawaguchi 1997, pp3 &7). This is similar the rates achieved by the German ICE train, without resorting to wheel/rail adhesion free braking methods.

Clearly there is significant scope for future improvement to braking performance, both from the perspective of braking force and adhesion utilisation. The absolute limits that would apply to the braking performance of different types of train are not currently known, but operational experience has shown that nominal rates of up to 1.5m/s^2 can be achieved on high speed trains,

$2.5m/s^2$ on rubber tyred metro trains and $3.0m/s^2$ on LRVs. It also seems reasonable to consider the limit of practical steel wheel / steel rail metro operation to be in the region of $1.7m/s^2$.

5.8 SUMMARY

The major limitation that applies to maximum achievable braking rates today is the adhesion available between the rail and wheel. As discussed in section 5.4.1, there appears to be significant scope for improvement in this area by increasing both the utilisation of available adhesion (through WSP systems, train design and driver information) and the consistency of adhesion management (through vegetation control, sanding, chemical augmentation, wheel/rail cleaning and track maintenance). However, improving wheel/rail adhesion and its utilisation are not the only method available for overcoming its limitations. Significant potential is also offered by wheel/rail adhesion independent braking systems.

The main obstacles to the application of wheel/rail adhesion independent braking systems in the UK would appear to be cost (of both installation and maintenance) and regulatory requirements. Whilst the issue of cost can be objectively assessed and compared with the financial and social benefits of the improvements available for any proposed application, the issue of regulation is more difficult.

The 'Regulation of Railways Act 1889' states that passenger trains in the UK must have continuous brakes that are self-applying in the event of any failure in the continuity of their action. This requirement would appear to preclude any brake system that is not fail-safe in application. The author has only come across one wheel/rail adhesion independent braking systems that would currently comply with 'fail-safe' requirements (the rubber tyred operation, as used on Paris Metro). However, it could be argued that electro-magnetic track brakes with a traction battery back-up (as used on Sunderland Direct) are fail safe in practice, since the battery is constantly topped-up when running. In the opinion of the author, an aerodynamic brake or permanent magnet track brake could also be made to comply with fail-safe principles.

It is interesting to note that, since simply applying a full brake application may not result in optimal stopping distances during conditions of poor adhesion, modern WSP systems are used during both service and emergency brake application. It is entirely possible for these systems to fail 'wrong side', holding the brake off and potentially failing to comply with the Regulation of Railways Act. In order to minimise the risk associated with this failure condition, WSP systems used for emergency brake applications are implemented on a localised (typically per bogie) basis. This approach could clearly also be adopted for other elements of the braking system, in order to utilise the higher braking performance that they offer where this can not be achieved in a fail-safe manner. This approach has already been adopted elsewhere in the world, where electro-magnetic and permanent magnet track brakes are used extensively, particularly as an emergency brake (Fulford et al. 2001, p87). As an example of this, the use of electro-magnetic track brakes is mandated for emergency braking of trains with a maximum speed above 145km/h in Germany (Schofield(2) 2000, p3).

Whilst this would not be allowed today under current UK regulations and standards, there would seem to be good potential for further developments that could make the use of these systems more acceptable and useful to most railway administrations – and perhaps justify a change in the UK standards. In particular, the author would propose three developments that appear to offer significant potential for future braking systems:

- Fail-safe application methods for wheel/rail adhesion independent systems, such as holding a permanent magnet track brake off by use of air or magnets and engaging it by means of springs;
- Control systems capable of adjusting the current applied to an electro-magnet's exciting coils in a fail safe manner, in order to maintain a consistent brake force at all speeds and;

- Localised brake control, so that non-fail-safe mechanisms can be used without failure of a component leading to total loss of braking capability (the approach used for WSP systems).

Whilst it seems likely that continued optimisation of braking systems will lead to higher braking rates being possible than have currently been demonstrated in operation, the author proposes to adopt values as the limits to be applied within his further research that are realistic. To this end, he will adopt assumed values that fall within the ranges discussed in sections 5.4.2 and 5.7. A summary of these values can be seen in Table 5-8.

Railway Type	Current Maximum Braking Rates (UK)		Assumed Limit of Deceleration Rates		Assumed Limit of Jerk Rates	
	Service	Emergency	Service	Emergency	Service	Emergency
High Speed Trains (Main Line)	$0.88m/s^2$ (nominal)	$1.17m/s^2$ (nominal)	$1.4m/s^2$ (nominal)	$1.5m/s^2$ (nominal)		
Low Speed Trains: Steel Wheel / Rail (metro and suburban commuter services)	$1.15m/s^2$ (nominal)	$1.1m/s^2$ (open); $1.3m/s^2$ (tunnel); (nominal) $1.98m/s^2 <$ $20km/h >$ $1.8m/s^2$ (Instantaneous)	$1.45m/s^2$ (nominal)	$1.7m/s^2$ (nominal)	$2.35m/s^3$	$9.0m/s^3$
Low Speed Trains: Rubber Tyred (metro services) & Light Rail Vehicles		$2.8m/s^2$ (nominal)		$2.8m/s^2$ (nominal)		

Table 5-8: Assumed Limits of Practical Deceleration and Jerk Rates

Whilst the author is happy to adopt these assumptions in the light of proven operational experience around the world, the true limits of adhesion or safe and comfortable deceleration and jerk rates for different types of train operation remain unknown. This is largely due to the low quantity and poor quality of available research. It is, therefore, possible that higher braking rates could be utilised if the effects of adhesion, acceleration and jerk were better understood. However, it is also possible that the adhesion available on a particular railway may be worse than that experienced elsewhere. The BR ATP scheme 'lessons learnt' seminar noted that "well founded assumptions have to be made for actual brake performance, rail-wheel adhesion etc." (Clementson 2001, p1).To this end, further fundamental research into the subjects of adhesion, acceleration and jerk (although outside of the scope of the author's work) is clearly required.

6 SPEED SUPERVISION CRITERIA

6.1 INTRODUCTION

Since ATP systems are provided as safety systems, they must ensure that unsafe levels of overspeeding do not occur. At the same time, in order to support efficient operation of the railway, they must not be overly restrictive in constraining a driver's normal (safe) driving behaviour.

If an ATP system is to prevent the train speed from reaching unsafe levels, it must:

- Allow for delays in brake application once an intervention has been made;
- Allow for the system's own processing delays;
- Allow for worst-case acceleration of the train during these delay periods.

It must, therefore, initiate an intervention at a speed well below the actual safe limit.

In addition to the constraints imposed by safety, many operators also require ATP systems to:

- Provide a warning to the driver if he/she is detected as overspeeding;
- Allow the driver some reaction time following a warning before intervention occurs;

If such facilities are provided, the speed at which warnings must be initiated will be even further below the actual safe limit. However, in order to ensure efficient operation, such warnings are generally required to commence only once overspeeding has occurred (not below the permissible speed).

Fortunately, the permissible speeds applied to the railway (whether line speeds, maximum train speeds or permanent, temporary or emergency speed restrictions) are not set at the maximum safe speed for travel. Indeed, during a lecture to the eighth IEE vacation school on Railway Signalling and Control Systems, it was noted that "generally speed restrictions are imposed either for comfort or to reduce wear and tear. There is usually a substantial margin for error. In the case of track based speed restrictions the overturning speed is typically 50% above the speed limit" (Fenner 2000, p1). Any such 'margin' between the permissible and safe speeds of travel for a train provides 'room for manoeuvre' in the development of ATP speed supervision criteria.

In order to identify the margins available to ATP speed supervision criteria, in this chapter the author considers the main reasons why speeds may need to be limited during operation and the methods used to determine permissible speeds and maximum train speeds on UK railways. The implications that this has for the design of ATP system speed supervision criteria will then be discussed.

A number of the reasons for setting speed restrictions relate to the alignment of the track – a topic not yet considered within this thesis. Therefore, in order to assist the reader's understanding, the author begins this chapter by briefly considering the subject of track alignment in order to define the terms that will be used in the subsequent discussion.

6.1.1 TRACK ALIGNMENT

In plan view, track consists of straight (tangent), curved and transitional sections, designed to maintain a constant track gauge (that is, a constant distance between the vertical faces of the running edges of the rails). The standard track gauge is 1435mm, giving a distance between rolling contact points (the actual running bands on the railhead) of 1500mm (Carney 2001, p16; Cope 1993, p245; Scott 2002).

Curves are generally defined by their radius and are 'handed', depending on the direction in which travel deviates. A right hand curve deviates to the right and a left-hand curve to the left. In a simple curve, two tangent sections of track will be joined together by a single circular curve of constant radius. As a train travels around such a curve, it is necessary to deflect the vehicles, which would naturally tend to continue in a straight line. This deflection is achieved as the result of centripetal (or lateral) forces acting inwards towards the centre of the curve, exerted on the train's wheelset by interaction with the track (Carney 2001, p14; Cope 1993, pp85-7).

In accordance with Newton's third law of motion, the track experiences a reaction that is equal and opposite to the centripetal forces that it is exerting on the wheelsets. This reaction, known as the centrifugal force[1], therefore appears to act horizontally outwards from the curve, with a magnitude equal (but opposite) to the centripetal force and given by:

$$Centrifugal\ Force = \frac{M V^2}{R}\ [N] \qquad \text{(Cope 1993, p97)} \qquad \text{Equation 6-1}$$

Where M is the vehicle mass (kg), V the speed (m/s) and R the radius of the curve (m).

If both rails of the track are horizontally aligned, objects within the vehicle will also experience deflection from the straight-line motion and the consequent centrifugal reaction to the curving forces. This reaction will cause the objects within the vehicle to tend to slide outwards. If one of the objects is a passenger, this will cause an uncomfortable sensation as if he/she were being pushed over. Cant is therefore applied to the track in order to reduce the effects of centrifugal forces (Cope 1993, pp87 & 97; Schmid[(2)] et al. 2002, pp17-18).

Cant is the difference in vertical level across the track gauge. It is measured in mm at the railhead centres or contact band of each rail, or as an angle between the plane of the track and the horizontal (see Figure **6-1**) (Blakeney[(1)] 2001, p8; Carney 2001, p16). If the track is canted inwards, the centrifugal force becomes resolved into two components, one perpendicular to the vehicle floor and the other parallel to it. Similarly, the forces produced by the mass of the train also become resolved into two components, one perpendicular to the vehicle floor and the other parallel to it in the opposite direction to the centrifugal forces (see Figure **6-1**). The two forces parallel to the vehicle floor therefore act against each other, reducing the effects of lateral acceleration on passenger comfort (Cope 1993, pp97-99).

As the forces due to the vehicle mass are constant but the centrifugal forces are dependent on the speed and curve radius, it is possible to design a curve with a level of cant that causes the perpendicular force components to cancel each other out at a given speed. This cant is referred to as the 'equilibrium cant' which, when achieved, produces no sensation of lateral force for passengers on the train (Carney 2001, p16; Scott 2002, p5).

---

[1] Whilst 'centrifugal force' is often viewed as a misnomer (since vehicles are simply trying to continue travelling in a straight line), it is a term widely used within the railway industry. This terminology has therefore also been adopted by the author. A fuller discussion of the term 'centrifugal force' can be found in BBC 2004.

Unfortunately, since the equilibrium cant can only exist at one speed on a particular curve, most trains will travel around the curve at a higher or lower speed. This causes a deficiency or excess of cant, in both cases leading passengers to experience the sensation of lateral forces. If the cant excess or deficiency becomes too high, the train may become liable to overturning – and long before that the limit of reasonable comfort for passengers and the driver will be reached (Cope 1993, pp97-99, 373). A balance must therefore be found in selecting suitable levels of cant and determining line speeds.

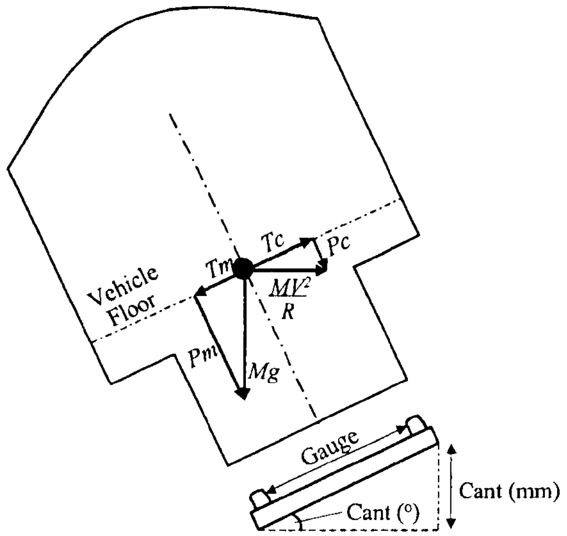


Figure 6-1: Weight and Centrifugal Forces on Canted Track

Cant deficiency and excess are measured by the angle (in degrees) or distance (in mm) that they differ from the equilibrium cant, rather than the speed differential that they represent.

If the track on a simple curve is canted and the tangent track to which it is joined is not, a cant gradient is introduced across the join. This can cause a sudden jerk in the train's movement, affecting both passenger comfort and safety. In order to avoid this occurrence, a transitional section can be included between the circular arc of the curve and the tangent track to gradually change both the horizontal curvature and track cant (Cope 1993, p374; Scott 2002, p5). 'The nature of the transition curve is probably the most important design feature of a railway alignment for high speed passenger vehicles as it has the greatest influence on passenger comfort' (Carney 2001, p14).

Horizontal curves are not the only element of track alignment to consider. Railways also require gradients in the track. As with horizontal transitions, vertical curves are used to minimise the transient effects of changes in acceleration where sections of track with different gradients join. They must limit the rate of change and value of vertical acceleration experienced by passengers to a comfortable value, limit wheel unloading during the transition, prevent excessive compression of the springs of the inner axles of any vehicles with three or more fixed axles and maintain track clearance (to avoid 'grounding') (Cope 1993, pp383-6).

Vertical curve transitions are usually required if the difference of the respective gradients (if they are of the same sense) or their sum (if they have opposite senses) is more than 2.5% (Profillidis 2000, p180). However, "in the context of rail speed increases vertical radii could be considered less important than horizontal curves" (Carney 2001, p15). They will not, therefore, be expanded upon here.

6.2 THE NEED FOR LIMITING SPEEDS

The speed of a rail vehicle can affect both the comfort and safety of passengers through a number of mechanisms, each of which raises the need for limiting speeds in certain circumstances, for certain vehicles or at certain locations. In order to consider the relevant mechanisms and their effects, the author has identified five reasons why speed may need to be limited, each of which will be considered in turn.

6.2.1 PASSENGER COMFORT

The basis for most track geometry and line speed limits is passenger comfort.

In France, various tests for quasi-static accelerations have enabled acceptable values for passenger comfort to be determined. These include:

- Vertical accelerations (due to track convexity up or downwards)
  - — Normal values $0.45 m/s^2$;
  - — Exceptional values $0.5 m/s^2$ due to track upwards or $0.6 m/s^2$ downwards;
- Lateral Accelerations (due to horizontal curves)
  - — Normal values $0.85 m/s^2$,
  - — Exceptional values $1.2 m/s^2$;
- Lateral Jerk (rate of change of lateral acceleration)
  - — Normal values $0.3 m/s^3$;
  - — Exceptional values $0.7 m/s^3$

  (Montagne 1975, p35).

Passengers are also subjected to random dynamic accelerations caused by track irregularities, which can cause physical weariness. Sensitivity to such irregularities has been found to be greater on curved than on tangent track and to reach a maximum at vibration frequencies in the order of 5 Hz (Carney 2001, p23; Montagne 1975, pp35-6; Profillidis 2000, p48). Studies have found that an acceleration of 0.05g at a frequency of 5Hz can be tolerated by the human body for up to 5hr 30min if it is in the vertical plane and for up to 3hrs 30 min in the horizontal plane. (Profillidis 2000, p171).

The effect of speed in increasing centrifugal forces has already been considered in section 6.1.1. In terms of static and quasi-static lateral forces, "it is the lack of cant or the excess of cant that is most felt by passengers" (Dyson 2000, pE11/3). A 110mm cant deficiency at the wheelset corresponds to $0.72 m/s^2$ lateral acceleration. Allowing for vehicle body roll in the suspension, this is generally increased to $1 m/s^2$ as felt by the passenger. For a tilting train running at 300mm cant deficiency with a tilt compensation of 75% (50% after allowing for body roll) the passenger still feels no more than $1 m/s^2$ (Dyson 2000, pE11/3). Hence, some of the adverse affects of lateral accelerations on passenger comfort can be overcome by operating tilting trains – enabling them to run comfortably around curves at higher speeds.

Another significant factor, contributing to the lateral jerk experienced by passengers, is the rate of change of cant (a measure of the rotation of the car body and the passengers within it). A European pre-norm suggests that a roll rate of $8^{\circ}/s$ makes less than 1 in 10 people feel uncomfortable. The current UK limiting value for conventional trains is $3.4^{\circ}/s$ and for tilting trains is $4.2^{\circ}/s$ (Dyson 2000, pE11/4).

In 1984 a high-speed train was run on a curved route at speeds up to 30% above normal curve speeds. At normal operating speeds, 12% of standing and 3% of seated passengers were dissatisfied with the ride comfort and that this dissatisfaction doubled when the line speed was increased by 10mph (Carney 2001, p22). "It has been shown that increasing speed reduces passenger comfort on transitions, on curves and at discrete events such as track defects and crossings" (Carney 2001, p32).

6.2.2 DERAILMENT

The derailment of a train can result from a number of causes. Those that can be speed related include:

- **Flange climb:** the flangeway clearance on a wheelset typically allows 7 to 10mm of lateral displacement to occur before the flange contacts with the rail. However, on tight curves this allowance may be insufficient to prevent the flange making contact with the rail. If the flange becomes unable to slide past the rail for a finite distance when this occurs (typically for a couple of metres), the wheel will climb up the rail, ultimately derailing the train. It is more likely that this will occur if there is a high lateral force between the wheel and rail, which can be caused by high speed (Clementson 2000, p6; Wickens 1998, p209);
- **Dynamic Reaction to track geometry:** if cyclic features in the track (such as dips in the rail at joints and welds) excite the suspension of freight wagons into resonance, they can start to bounce violently – possibly right off the rail. Where cyclic conditions occur, the speed of vehicles needs to be restricted to ensure they can not reach their natural frequencies (which will be dependent on both vehicle type and loading – occurring at lower speeds for fully loaded vehicles, but more violently for unloaded vehicles) (Clementson p9-p10; Lewis 2000, p5);
- **Gauge Spreading:** on curved track with relatively high lateral curving forces the wheels can move closer together on an axle or the gauge can become excessively widened. If either effect occurs, the wheels drop off the rail and into the four-foot. Gauge spread develops over time and the speed of an individual train is not generally a significant factor, but the lateral forces exerted by all trains increase in proportion to the square of the speed, making gauge spreading more likely on lines with higher permissible speeds (Clementson 2000, p15; Dyson 2000, pE11/1-2);
- **Track Shifting:** If the transverse forces generated by a passing vehicle exceed the track transverse resistance the track will shift, which can cause derailment. The transverse track resistance depends on the type of sleepers and track maintenance used. It is highest on concrete sleepers with mechanical means of maintenance and lowest on track with timber sleepers and manual maintenance. In all cases, it is lowest just after track maintenance, which destabilises the track. Transverse forces are produced by unbalanced centrifugal acceleration on curves as well as the dynamic effects of track and rolling stock defects. This form of derailment is, therefore, more likely at high speeds (Profillidis 2000, pp156-7, 164).

Where there is significant risk of any of these derailment modes occurring, speeds may need to be limited – either on a permanent or temporary basis. It would also be standard practice to impose an emergency speed restriction in the case of cyclic irregularities or gauge spreading becoming apparent.

6.2.3 *OVERTURNING*

A train subjected to high lateral forces is at risk of overturning. This can occur under three conditions:

- At high speeds on curved track a combination of overspeed, curve misalignment, cant loss and cross wind can result in excessive cant deficiency and unloading of the inner rail;
- Light weight empty wagons standing still or with a displaced freight load can experience unloading of the outer rail on track with high cant and an unfavourable cross wind;
- Unloading of the outer rail can also occur when a train failure results in non-released brakes at the rear of the train.

  (Carney 2001, p33).

In general, the main track related factors influencing the overturning of a train are the design cant and curve radius. Irregularities in cant and curve radius only become significant if they extend for the length of the train (Blakeney[(1)] 2001, p14; Cheesewright 1999, p2).

Rail group standards in the UK require vehicles to be designed with mass distribution and suspension characteristics that allow them to travel around smooth curves at constant speed without rolling over at not less than 19.5° cant deficiency for freight vehicles designed to operate at up to 75mph and not less than 21° cant deficiency for all other vehicles (Blakeney 2000, p6).

When it is considered that the maximum cant deficiency permitted for a conventional train on UK main line infrastructure is 150mm (5.74°), it can be seen that there is, at worst, a margin of 15° cant deficiency between the train's permissible speed and its roll-over resistance. This margin is reduced to 9° cant deficiency for tilting trains, which are permitted to operate at cant deficiencies of up to 300mm (11.5°) (Blakeney[(1)] 2001, p5; Carney 2001, p52; Cope 1993, p451).

A train travelling at a nominal design cant deficiency may experience significantly higher cant deficiency in practice, due to the effects of imperfect wheel condition, track irregularities (especially in cant and curve radius), wind and possible overspeeding of the train (Blakeney[(1)] 2001, p5; Dyson 2000, pE11/1). Factors such as these reduce the effective margin to overturning and, therefore, must be accounted for in the design of appropriate track, trains and speed limits.

The levels of track irregularities experienced on UK main line railways has been measured by track recording coaches. From the measurements taken, it has been concluded that for a design curve radius of less than 3000m, the minimum curve radius expected is 80% of the design radius. For curves with radii of 3000m or greater, the minimum expected is 70% of design radius. In most cases, deviation in cant is less than 10mm (0.38° cant deficiency) (Cheesewright 1999, p3). Therefore, fixed allowances can be made for these factors in determining speed restrictions. However, if all of the effects that could lead to a train overturning were considered to act simultaneously, the permissible speed that could be considered acceptable would be very restrictive. Therefore, probabilities of wind speed and overspeed, which represent the most significant factors in overturning risk, are considered, rather than absolute values (Cheesewright 1999, piii & 16). To this end, the calculation of permissible speeds on UK main lines includes:

- Fixed allowances for reduction in curve radius and loss of cant;
- Probabilistic allowances for overspeed and the effects of wind;
- A fixed safety margin of 2° cant (for inaccuracies in the calculation of other allowances).

An example of how this is done is given in section 6.3.

In the development of permissible speed calculations, the tolerable probability of overturning per 100km of route section is considered to be 4×10^{-9}, including an average tolerable probability of overturning per curve of 4×10^{-11} and a maximum tolerable probability of overturning per curve of 10^{-9} (Blakeney[1] 2001, p24).

6.2.4 PERMISSIBLE FORCES

As a train travels along the railway it exerts vertical, lateral and longitudinal forces on the track and its support structures, causing fatigue in their materials and deterioration of the track geometry.

The longitudinal forces on the rail result from acceleration and deceleration during train operation. They are dependent on both train design and operating speeds, and must be considered in the setting of speed limits and restrictions so as not to generate forces which could damage the structural integrity of the rails, track and infrastructure. Longitudinal forces are particularly taken into account in the design of bridges on railway lines (Boocock 1993, p10; Profillidis 2000p47).

When subjected to vertical forces, the behaviour of the rails and sleepers is elastic, while that of the ballast and the subgrade is elastoplastic (Profillidis 2000, p46). The vertical forces produced by the passage of a train consist of four components:

- The normal static load of the wheel;
- The quasi-static load resulting from cant deficiency in curves;
- The random (dynamic) overload caused by rail to wheel interaction of un-sprung mass when the train runs over variations in the vertical alignment of the rail (high frequency 20-150Hz);
- The random (dynamic) overload caused by rail to wheel interaction of sprung mass (low frequency 0-20Hz)

  (Montagne 1975,p36).

All four components are influenced by the train's design, whether directly by the mass concerned or as a result of its suspension arrangements. The quasi-static and dynamic forces are also influenced by track design, track quality and train speed.

Rail Group standards in the UK require vehicles to be capable of negotiating a vertical ramp equivalent to a dipped rail joint on tangent track at their maximum design operating speed without exceeding a total vertical force (static and dynamic) of 322kN per wheel(Boocock 1993, p7). Vertical forces must also be considered in the setting of speed limits and restrictions so that trains will not generate forces that could damage the structural integrity of the rails, track and infrastructure (including bridges and viaducts) over the normal range of vertical track irregularities at normal operating speeds.

Just as with longitudinal forces, the lateral forces on the rail also result from several causes:

- Inclination of the wheel / rail contact patch (occurring particularly where wheel conicity is high – whether by design or following wear);
- Creep forces across the contact patch;
- Dynamic instability (hunting);
- Lateral curving (centrifugal) forces;
- Inertial forces generated at track irregularities.

  (Carney 2001, p41; Dyson 2000, pE11/2).

Railway Group standards require vehicles to be designed so as not to generate lateral forces that could jeopardise the structural integrity of the rails and track. The maximum permitted lateral force is given by:

$((M / 3) + 10)$ kN, Where M is the static load (Boocock 1993, p8) Equation 6-2

This maximum applies if the force is sustained over 2m or more. A larger force acting over a shorter distance is considered acceptable because the risk of the track having time to move in response to the force becomes remote (Dyson 2000, ppE11/1-2).

Vehicles must also be capable of negotiating a lateral ramp discontinuity in track alignment when travelling on a curve at maximum normal operating speed and cant deficiency without exceeding a lateral force of 71kN (Boocock 1993, p9).

In the UK, conventional trains are not permitted to operate at more than 150mm cant deficiency and the static and quasi static forces that occur when operating under these conditions account for only 25% of the lateral force limits. For tilting trains, which can operate at up to 300mm cant deficiency, the static and quasi-static forces account for 50% of the lateral force limits (Dyson 2000, pE11/2).

The dynamic component of the lateral forces can be controlled by maintaining track to sufficient quality, designing trains with sufficient suspension and by limiting train speeds, since lateral forces increase in proportion to the square of the speed (Dyson 2000, pE11/1-2).

6.2.5 TRAIN SEPARATION / BRAKING DISTANCES

The final need for limiting train speeds that will be considered in this chapter is the requirement to keep trains separated by safe braking distances.

Some sections of line (particularly straight plain line sections) are physically capable of safely supporting far higher operating speeds than are actually permitted over them. Where this is the case, it would be physically possible to set a higher permitted speed. However, for a section of line with an existing signalling layout, the line speed can only be increased for trains capable of sufficient braking performance to stop the train from the higher speed within the distances required by the signal spacing – otherwise safe train separations can not be guaranteed. It is likely, therefore, that any increase in permitted speeds would need to be accompanied by an improvement in braking performance.

For a new section of line (or one undergoing resignalling), it would be possible to design the new signalling to support any desired speed of operation, whilst maintaining adequate separation between trains. However, the need for longer braking distances from the higher speeds would result in a reduction in the headway that could be achieved on that section as the permitted speeds were increased. The train headway required to deliver the desired capacity on the line, the train's braking performance and the design of any signalling layout are all factors that must, therefore, also be considered in setting permissible speeds for a section of line.

6.3 SPEED SUPERVISION

The permitted speeds in use on British main line railways have generally been set on the basis of assumed manual driving without speed supervision. As a consequence of this, some level of overspeeding has been assumed in setting them.

According to Rail Group standards, train-operating companies are required to carry out checks on the speeds of their trains. In so doing, any incidence of the permissible speed being exceeded by more than 3mph must be reported to Network Rail as quickly as possible. The action to be taken beyond this depends on the level of overspeed:

- If the permissible speed is exceeded by 4 or 5 mph (6 to 8km/h), the driver must be informed and the facts recorded;
- If exceeded by 6 to 10mph (10 to 16km/h), the driver must be informed and interviewed at the first scheduled stopping point where a competent person is available to establish both the facts and the driver's fitness to continue duty;
- If exceeded by 11mph (18km/h) or more, Network Rail must arrange for the train to be stopped at the first available point where the driver can be informed of the infringement and interviewed by a competent person to establish both the facts and the driver's fitness to continue duty

  (Evans 1995, p5).

Some explanation of the criteria used for responding to driver overspeeding can be found in GE/RC 8517 (recommendations for systems for the supervision of enhanced permissible speeds and tilt enable), where it is noted that:

- A mechanical speedometer typically requires a tolerance for driver control of 4km/h;
- Below 160km/h, speed measurement precision is typically 6km/h.

  (Blakeney 2001[(2)], p33)

A driver's real speed could, therefore, be 10km/h above the permissible speed before it is reasonable to assume that he/she is overspeeding.

With the introduction of ATP, it becomes necessary to consider the speed supervision criteria that should be used in determining warning and intervention limits onboard the train. These must be selected so as to avoid intervention when the driver is correctly controlling the train's speed, as well as to minimise the potential for overspeeding. In doing this, tolerances including the accuracy of speed measurement (such as wheel diameter calibration, slip and slide, sensor accuracy) and the dynamics of controlling the train's speed (such as delays in traction cut off and brake application) must be accounted for.

As an example of this, if the driver of a class 390 realised that his/her train was exceeding the permitted speed the train's speed could continue to rise by 9.3km/h before he/she would be able to bring the train's speed under control under worst case conditions. This figure is based on:

- 2 second allowance for driver reaction;
- 1 second delay in traction cut off;

- Maximum 3% down gradient[2];
- 4 second brake build up delay (consisting of 2 second delay to start of application, followed by assumed linear build up);
- Maximum acceleration 0.44m/s^2 below 70km/h, then decreasing linearly to 0.07m/s^2 at 230km/h;
- Brake Rate 9%g;
- 1.6km/h allowance for speed measurement error (based on a specified maximum speed measurement error of ±0.5km/h ±0.5% of speed, with maximum speed of 225km/h);

  (Barnard[(1)] 2002, p31; ERTMS Users Group[(1)] 2002, pp8-11; Fleming 2000, p14)

If automatic speed supervision is configured to provide a warning prior to brake intervention, the margin between the warning and brake intervention must, therefore, be at least 9.3km/h if the driver is to be allowed to act on the warning without invoking an intervention under all circumstances. This is typical of other train classes, the margin between warning and intervention generally needing to be between 5 and 10km/h (Blakeney 2001[(2)], p38). At higher speeds, when trains tend to have poorer acceleration performance, the required margin is smaller. Returning to the example of a class 390, the maximum achievable acceleration at 200km/h is only 0.07m/s^2. This reduces the required margin between warning and brake intervention to around 4km/h.

If the warning is to be given only once the train is definitely travelling above the permitted speed, a further margin of 5.5km/h would be required for a class 390 (allowing 4km/h tolerance for driver control and 1.5km/h for possible speed measurement error). The train could, therefore, achieve a speed nearly 15km/h above the permitted speed if the driver applied the brakes on receipt of an overspeed warning.

6.3.1 EXISTING ATP SPEED SUPERVISION CRITERIA

The first ATP systems to be installed on the UK main line railways were the two BR-ATP prototype projects. The supervision criteria used for these included:

- 3mph (5km/h) overspeed when comparing the system's speed measurements with the permissible speed before providing an audible warning to the driver (implying between 3.4 and 6.6km/h overspeed when comparing actual speed with the permissible speed);
- 6mph (10km/h) overspeed when comparing the system's speed measurements with the permissible speed before an automatic service brake application (implying between 8.4 and 11.6km/h actual overspeed);

  (Holgate et al 1993, p2).

For a class 390, the 5km/h margin between permitted and warning speed allows the driver a 3.4km/h tolerance for controlling the train to the permitted speed under worst-case conditions. Whilst this is slightly lower than that already stated as typically required for a mechanical

---

[2] Whilst the Railtrack standard for 'Lineside Signal Spacing' includes separations for up to 3%g, the author is not aware of any gradient steeper than 2.5%g on the UK rail network (RLE 2002). However, elsewhere in Europe such steep gradients do occur - for example, on the TGV Sud-Est (Schmid[(2)] et al. 2002, p2). Since any future ATP system is likely to be utilised throughout Europe and must be able to cope with the worst case conditions that it will encounter, the author has chosen to assume a 3%g maximum gradient.

speedometer, it has proved successful in operation. The further 5km/h margin between warning and service brake intervention would be sufficient for driver reaction to prevent an intervention on a 3% down gradient, with just over $0.05m/s^2$ traction acceleration. It would also permit initial traction acceleration of $0.44m/s^2$ on gradients of up to 0.15% down. Reducing the margin to 4km/h would only guarantee sufficient margin between warning and service brake intervention for gradients of 2.6% down or less with no traction acceleration. The BR-ATP margins therefore appear to be almost at the ideal level for the class 390 - if it is assumed that drivers will not be using traction acceleration once their train has passed above the permitted speed and that the Class 43, 165 and 168 trains operating under BR-ATP have similar traction cut off and brake build up performance to the class 390.

Railtrack specified the later Tilt Authorisation and Speed Supervision system (TASS)[3] to supervise ceiling speeds against train specific margins. These are calculated based on the train's traction cut off and brake application delays, such that:

- For the class Class 390, a warning margin of 4.5mph (7.2km/h) and an emergency brake intervention margin of 6.1mph (9.8km/h) above the supervised EPS are required;
- For the class 221, a warning margin of 3.5mph (5.6km/h), service brake intervention margin of 5.5mph (8.9km/h) and emergency brake intervention margin of 6.5mph (10.5km/h) are required.

  (Barnard[(1)] 2002, pp35 & 42)

It is worth noting that TASS does not give the driver an indication of the supervised speed within his/her cab. This fact explains the larger margin provided between the EPS and warning speeds when compared with the 5km/h margin for BR-ATP, which provides an indication of the permitted speed. Whilst the TASS margins between permitted and warning speeds should be ample to allow driver flexibility, the small margin selected between warning and intervention means that it may not always be possible for the driver to respond to a warning quickly enough to prevent a subsequent system intervention. The use of such small margins is, in part, based on an assumption that drivers should not be accelerating at full rate as they closely approach the then current permissible speed. However, this explanation alone does not fully justify such small margins, since they would only be sufficient to permit the driver 2 seconds reaction time in applying the brake following a warning if:

- No traction power was being applied and;
- The train was on an average gradient of 1.7% down or less during that interval.

When the author raised this with the TASS design team prior to the system's implementation, it was accepted that defined margins might be inadequate. However, no difficulties were encountered following implementation due to the professionalism shown by drivers in avoiding TASS warnings (see Appendix A, interview with Bob Barnard, question 2).

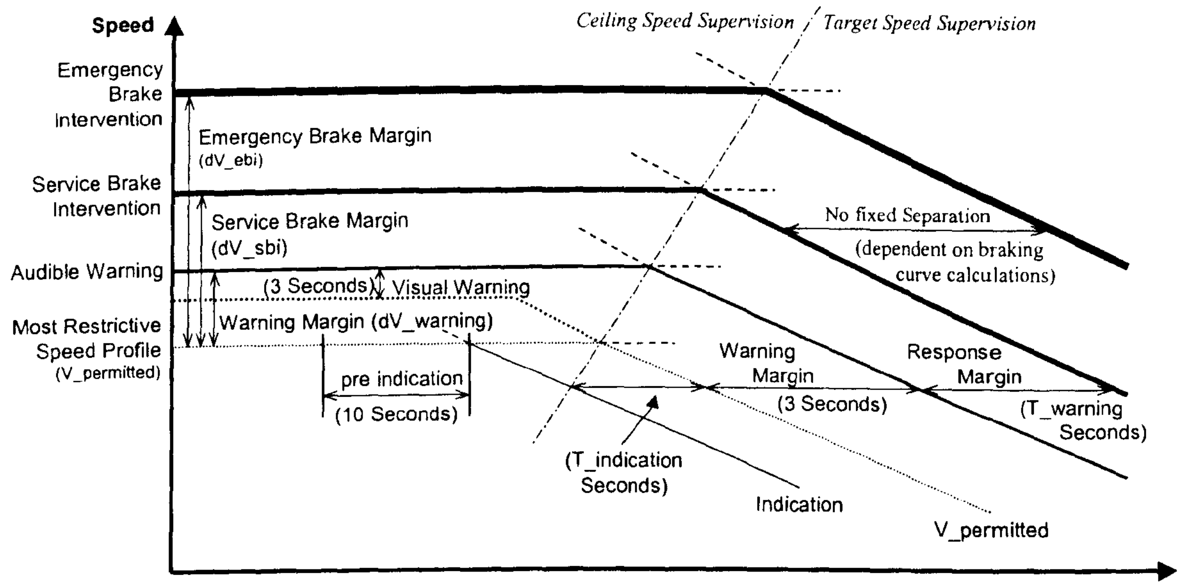
Another system that is currently being developed for implementation on UK main line railways is the ETCS ATP system (both levels 1 and 2). As this is a standard European system, the margins to be used in its speed supervision are being defined by the ERTMS Users Group. Railtrack lobbied within this group for the ceiling speed supervision criteria to give the driver an audible warning at 5km/h overspeed, a service brake intervention at 10km/h overspeed and

[3] TASS – a system for supervising the speed of tilting trains through areas of Enhanced Permissible Speeds

an emergency brake intervention at 15km/h overspeed (ERTMS Users Group 2001, pp7-10). Like BR-ATP, the ETCS systems will provide the driver with an in-cab indication of the permitted speed. The warning and service brake margins within this proposal are identical to those used in BR-ATP, and their suitability has already been discussed. If a class 390 is considered once again, the train's speed could increase by up to 4.1km/h following a service brake intervention, if the train were accelerating with full traction power on a 3% down gradient at the moment of intervention. The proposed 5km/h margin between service and emergency brake intervention would, therefore, appear to be sufficient to guarantee that emergency brake intervention will only occur if the service brake fails to apply.

The supervision criteria adopted by the ERTMS Users Group were not ultimately in accordance with Railtrack's proposal, opting for speed dependent, rather than fixed, margins. Despite Railtrack's concerns that the superior acceleration capability of modern trains increases the risk of overspeeding at low speeds, the criteria to be used in the ETCS systems will be as shown in Figure 6-2 (ERTMS Users Group 2001, pp7-10; ERTMS Users Group[1] 2002, pp8-11).

It should be noted that the ETCS emergency brake intervention curve is calculated on the basis of 'guaranteed deceleration' (assuming the worst case conditions for available brake performance, traction and gradients). It also has a target stopping point at the 'supervised location' (the end of the signal overlap or the point of danger if no overlap is available). In contrast to this, the service brake intervention curve is calculated on the basis of 'expected deceleration' rates with a target stopping point at the signal being approached (and a second objective of avoiding emergency brake intervention, if that requires more restrictive supervision) (ERTMS Users Group[1] 2002, pp8-9)



dV_ebi has a minimum value of 3 km/h, ramping up linearly to 15 km/h at 200 km/h and then remaining at that level;
dV_sbi has a minimum value of 2.5 km/h, ramping up linearly to 12.5 km/h at 200 km/h and then remaining at that level;
dV_warning has a minimum value of 2 km/h, ramping up linearly to 8 km/h at 200 km/h and then remaining at that level;
T_warning is equal to the traction cut off delay, brake build up delay and a 2 second allowance for driver reaction
T_indication is equal to the traction cut off delay and brake build up delay

Figure 6-2: ETCS Speed Supervision Curves (based on ERTMS Users Group[1] 2002, pp8-11; ERTMS Users Group[2] 2002, p18)

The small margins selected between warning and intervention for ceiling speed supervision in the ETCS system mean that it may not always be possible for the driver to respond to a warning quickly enough to prevent a subsequent system intervention:

- The margin between warning and service brake intervention would only permit the driver of a class 390 2 seconds of reaction time in applying the brake following a warning at 80km/h if no traction power was being applied and the train was on an average gradient of 1.4% down or less during that interval;
- The margin between permitted speed and service brake intervention mean that if a class 390 reached the permitted speed without traction or braking applied, it would not be possible to avoid exceeding a service brake intervention on a 3% down gradient at speeds below 42km/h. At a permitted speed of 5km/h, it would only be possible to prevent service brake intervention under these conditions if the gradient was no steeper than 1.85%;

From the driver's perspective, the likelihood of intervention following a warning is mitigated to some extent by the fact that service brake intervention will only be maintained until the train's speed comes back below the permissible speed. However, this still seems likely to provoke irritation in any driver affected by such an intervention. Whilst at high speeds the emergency brake intervention margin is sufficiently above the warning margin to ensure that the driver (and ATP system) have time to bring the train's speed under control prior to an emergency brake intervention, this does not appear to be the case at low speeds.

As some examples of this, assuming operation of a class 390 with driver reaction delays of 2 seconds, no ETCS system reaction delays (a very generous assumption) and no traction power being applied to the train during reaction and brake build up delays:

- If the driver did not respond to a warning on a 3% down gradient, a service brake intervention would only be able to bring the speed under control quickly enough to prevent an emergency brake application at speeds in excess of 197km/h. At 80km/h, the emergency brake intervention could not be prevented on a gradient of 1.8% down or more, whilst at 5km/h it could not be prevented on a gradient steeper than 0.78% down;
- Even if the driver did respond to the warning after a 2 second delay, an emergency brake intervention would snot be prevented for a permitted speed of 114km/h (warning speed of 120km/h) or less on a 3% down gradient. At a permitted speed of 80km/h it could also not be prevented if the gradient was steeper than 2.3% down or if steeper than 0.84% down at 5km/h;
- The margins between permitted speed and emergency brake intervention mean that if the train reached the permitted speed without traction or braking applied, it would not be possible to prevent it exceeding the emergency brake intervention speed on a 3% down gradient at speeds below 26.5km/h. At a permitted speed of 5km/h, it would only be possible to prevent emergency brake intervention under these conditions if the gradient was no steeper than 2.2%.

Clearly, when the possibilities of traction acceleration and ETCS reaction delays are considered, these situations become even worse. It is currently unclear how these problems are to be overcome, since Railtrack's objections have now been overruled (ERTMS Users Group 2001, pp7-10; ERTMS Users Group[1] 2002, pp8-11).

6.3.2 IDEAL ATP SPEED SUPERVISION CRITERIA

The purpose of any ATP system is to protect against error by the driver (whether human or automated). It should not cause undue restrictions to the normal (safe) operation of the railway.

To that end, an ideal ATP system should never prevent the driver form properly performing his/her function in accordance with his/her training (by warning or intervening unnecessarily). It should also only initiate irrevocable brake applications (e.g., the emergency brake) as a last resort. This means that:

- In order to ensure that drivers can confidently drive at the permitted speed without fear of experiencing a warning or intervention, the ATP system must allow some margin for fluctuations around the permitted speed and must never warn or intervene when the speed indicated on the speedometer is below the permitted speed;
- It must always be possible for the driver to prevent an ATP brake intervention by recognising that the train is at the permitted speed and applying the brake to prevent the speed continuing to rise above the intervention threshold;
- Any driver who continues accelerating above the permitted speed has made an error that could cause significant overspeed. In this situation, it would seem acceptable for a service brake intervention to be initiated if the train's speed exceeds an intervention threshold. In such circumstances, it should be possible to release the brake intervention once the train's speed has returned to the permitted speed or below;
- If the train speed drifts above the permitted speed due to gravitational effects and the driver is given a warning by the ATP system, the margin between warning and intervention must always be sufficient to allow the driver to prevent an ATP brake intervention;
- If the ATP system initiates a service brake application, the margins allowed between service and emergency brake intervention must be sufficient to ensure that correct activation of the service brake is able to prevent an emergency brake intervention.

These requirements can be used to comprehensively define an ideal ATP system's warning and intervention margins:

1. The warning margin must always be at least the system's maximum speed measurement error above the required fluctuation margin;
2. The margin between permitted speed and service brake intervention must always be sufficient to allow a 2 second driver reaction delay, followed by traction cut off, brake application and build up delays, assuming a 3% down gradient and maximum traction acceleration;
3. The margin between warning and service brake intervention must always be sufficient to allow a 2 second driver reaction delay, followed by brake application and build up delays, assuming a 3% down gradient;
4. The margin between service and emergency brake intervention must always be sufficient to allow for traction cut off, brake application and build up delays, assuming a 3% down gradient and maximum traction acceleration without triggering an emergency brake intervention.

Clearly, these criteria are dependent on the performance of any particular train type. Generic margins could be determined on the basis of the worst case conditions for all train types, but these would not permit optimal performance.

Assuming that the required fluctuation margin is 4km/h (Blakeney 2001[2], p33), the margins required by these criteria for a class 390 would be at least:

- Warning Margin:

  4.6km/h + 0.005V, where V is the train's speed in km/h Equation 6-3;

- Margin between warning and service brake intervention:

  $$((0.03 \times 9.81) \times 4.\dot{3} = 1.2753 m/s = 4.6 km/h \qquad \text{Equation 6-4;}$$

- Service brake intervention margin will be the worst of 4.6km/h above the warning margin and:

  1.2753 + (3 × Amax) m/s Equation 6-5

  Since the acceleration performance of a class 390 is given by:

  Amax = 0.6 – 0.00832V m/s^2, where V is >= 20m/s

  = 0.44 m/s^2, where V is < 20m/s (Barnard[1] 2002, p32) Equation 6-6

  This gives a minimum margin between permitted speed and service brake intervention on a class 390 of:

  9.2km/h + 0.005V, where V is > 28km/h (based on Equation 6.3) Equation 6-7

  9.34 km/h where V is < = 28km/h (based on Equation 6.5) Equation 6-8;

- Margin between service and emergency brake intervention:

  $$((0.03 \times 9.81) \times 2.\dot{3}) + A\max \ \ m/s \qquad \text{Equation 6-9}$$

  That is:

  4.63 – 0.00832V, where V is >= 72km/h

  4.06 km/h where V is < 72km/h Equation 6-10;

- Margin between permitted speed and emergency brake intervention:

  13.83 – 0.00332V, where V is >= 72km/h (based on Equation 6.7)

  13.26 + 0.005V km/h where 72km/h >V > 28km/h (based on Equation 6.7)

  13.4 km/h where V is < = 28km/h (based on Equation 6.8) Equation 6-11

These margins are represented in Figure 6-3.

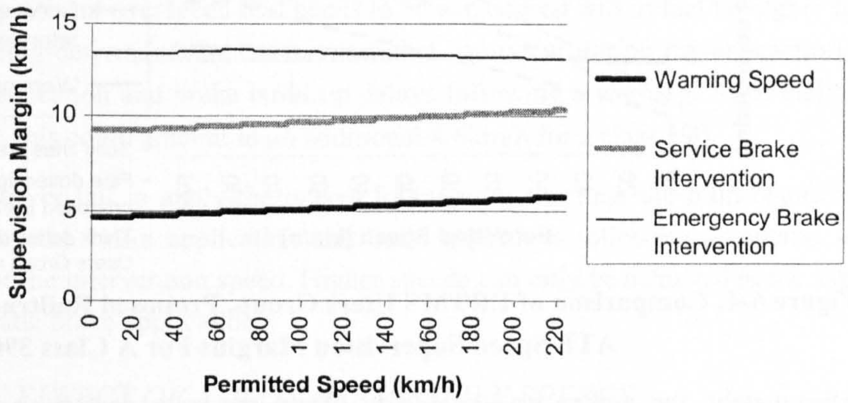


Figure 6-3: Ideal ATP Speed Supervision Margins For A Class 390

Assuming a maximum permissible speed of 225km/h, if flat rate rather than speed dependent margins were required for a class 390, the speed supervision criteria outlined in this section would require:

- Warning at least 5.8km/h above permitted speed in accordance with the system's speed measurements;
- Service brake intervention at least 10.4km/h above permitted speed in accordance with the system's speed measurements;
- Emergency brake intervention at least 13.7km/h above permitted speed in accordance with the system's speed measurements.

These values are remarkably close to those used in the BR-ATP systems and proposed by Railtrack for the future ERTMS warning and service brake intervention margins. However, if the assumption of drivers requiring an allowance of 4km/h fluctuation in order to control a train's speed about the permitted speed is true, it would appear that these margins should in fact both be increased by about 1km/h. Since the BR-ATP system margins have now been proved through successful operation, this would imply that a smaller allowance for fluctuation (in the region of 3km/h) is actually acceptable. This would be in keeping with the views expressed by Paul Le Vesconte (see transcript of interview, Qu11, in Appendix A). An alternative explanation for the difference could be derived from variation in performance (such as brake application and traction cut off delays) between the class 390 and the trains operated under BR-ATP (the class 43, 165 and 168). Unfortunately, this explanation cannot be tested in the absence of performance data for those classes of train (which the author has been unable to obtain).

In contrast to this, it would appear that the Railtrack proposal for ERTMS emergency brake intervention margins was higher than necessary. As this could have knock on effects on the safe level of enhanced permitted speeds, a reduction of 1km/h on the proposed margin would seem beneficial (possibly more if less than 4km/h fluctuation allowance is required).

What is plainly clear is that the margins adopted by the ERTMS Users Group are significantly inadequate at low speeds (see Figure 6-4). Since 'the smaller the tolerances allowed, the more difficult it is to avoid unwarranted interventions by the system' (Blakeney 2002, p21), they are likely to have a significant detrimental effect on operational performance if implemented as currently planned.

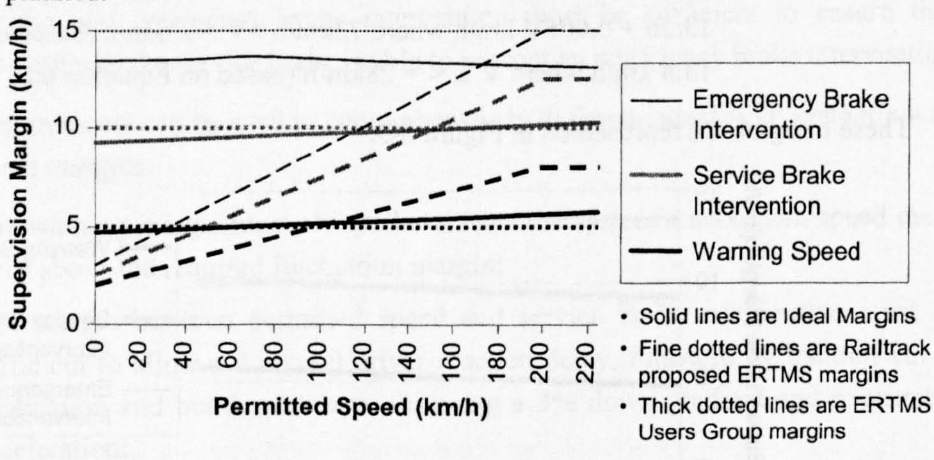


Figure 6-4: Comparison of ERTMS Users Group, Proposed Railtrack ERTMS and Ideal ATP Speed Supervision Margins For A Class 390

Unfortunately, the author was unable to obtain the train performance data that would be required to calculate the speed supervision margins of train classes other than the 390. He has, therefore, not found it possible to define a more generic set of supervision margins that could be applied to all train classes.

6.3.3 PROBABILITY OF OVERSPEEDING

The probability of overspeeding by different amounts has been determined for trains with no speed supervision by analysis of speed check records. The results of this analysis can be seen in Table 6-1, along with predicted probabilities for equivalent levels of overspeed with a Speed Supervision System using 5km/h Warning and 10km/h Intervention Margins calculated by AEA Technology Rail.

Overspeed (km/h)	Probability with No Speed Supervision	Probability with a Speed Supervision System using 5km/h Warning and 10km/h Intervention Margins
None	0.9	0.9
0 – 5	0.082	0.082
5-10	0.015	$1.5E^{-05}$
10-15	0.003	$2.6E^{-06}$
15-20	$4.6E^{-04}$	$4.6E^{-07}$
20-25	$8.2E^{-05}$	$8.2E^{-08}$
25-30	$1.5E^{-05}$	$1.5E^{-08}$
30-35	$2.6E^{-06}$	$2.6E^{-09}$
35-40	$4.6E^{-07}$	$4.6E^{-10}$
40-45	$8.2E^{-08}$	$8.2E^{-11}$
45-50	$1.5E^{-08}$	$1.5E^{-11}$

Table 6-1: Probability of Attaining Levels of Overspeed (based on Blakeney 2001, p43; Cheesewright 1999, p71 & 74)

The overspeed probabilities in Table 6-1 assume a 0.5 second delay to brake application and 3 second brake build up, along with constant speed during these delays (Cheesewright 1999, p67). The warning and intervention limits represented are consistent with those used with BR-ATP and similar to those adopted for TASS. They are also roughly in line with those identified in section 6.3.2 as being ideal for a class 390. The ramped characteristic of the ETCS supervision criteria would, however, suggest probable levels of overspeed that would be lower at low speeds and higher at high speeds.

The analysis conducted in determining the probabilities shown in Table 6-1 concluded that the overturning risk resulting from a driver failing to apply the brakes at the speed supervision system's warning margin is negligible (Cheesewright 1999, pp17 & 67). Since, in contrast to the assumption made in calculating the figures within Table 6-1, the train could accelerate (due to the effects of traction power and gradients) even if the driver responds to a warning, the maximum significant overspeed that needs to be considered will in fact be higher than this. The increment being determined by the increase that can occur during driver reaction, traction cut off, brake application and brake build up delays following a warning. As already discussed in section 6.3.2, this could amount to an additional 4.6km/h for a class 390.

Should the driver fail to apply the brakes following a warning, the train could reach a speed determined by the brake application and build up delays following automatic service brake application at the intervention speed. Higher speeds can only be achieved in the event of failure of the automatic brake application.

6.3.4 THE EFFECT OF ATP ON PERMISSIBLE SPEEDS

In section 6.2 the author discussed the main reasons that speeds may need to be limited on a railway. The two most significant of these factors are passenger comfort and overturning risk.

For a conventional (non-tilting) train, the constraints of passenger comfort mean that large margins are available between permissible and overturning speeds during normal operation. Since ATP can do nothing to improve passenger comfort, it appears that the introduction of ATP can have no effect on the setting of permissible speeds (it can only ensure that the speeds determined as acceptable for passenger comfort are not exceeded by too large a margin). However, this is not quite true.

Speed restrictions on UK main line railways are always quoted in 5mph (8km/h) intervals (Dyson 2000, pE11/4). This is done to assist drivers in remembering the permitted speed for the section of track they are travelling over. Whilst trains are driven on the basis of route knowledge, or on the basis of limited aspect speed signalling through lineside signals, this is a necessary simplification.

Since the permissible speed must always be selected as a safe and comfortable speed, it must always be rounded down to the nearest 5mph (8km/h) interval. This small, but finite, reduction will have a minimal effect on journey times at high speed (up to 3.5% at 200km/h), but could be significant at lower speeds (for example, up to 14% at 50km/h).

If a railway were to be operated on the basis of in-cab signalling that includes an indication of the permitted speed, it could be possible to refine the permitted speeds to the nearest 1mph (or 1km/h if the railway were to switch to the standard European units). However, in the opinion of the author, the methods used for determining speed restrictions are not currently precise enough to make full use of this potential refinement. It is, therefore, unlikely that any significant benefits could be gained by attempts to implement it.

Tilting trains can travel at enhanced permissible speeds (EPSs), much closer to their overturning speed than conventional trains can reach without causing undue discomfort to passengers (Metcalf 1999, p3). This often makes the overturning risk the main constraint on tilting train speed, rather than passenger comfort. In a similar way, temporary speed restrictions (TSRs) are sometimes set to mitigate the risks of derailment or overturning in areas of poor infrastructure. Here ATP can have a significant effect on determining safe EPS and TSR values. By reducing the probability of any overspeed, the ATP system reduces the risk of the train attaining overspeed levels that could result in overturning or derailment. The enhanced permissible speeds that can be considered safe for a tilting train and some TSRs for all trains will, therefore, become directly related to the ATP's supervision criteria implementation. If the probable overspeed can be reduced, the EPS/TSR can be increased by an equivalent amount without increasing the overturning or derailment risk (see Figure 6-5 for an example).

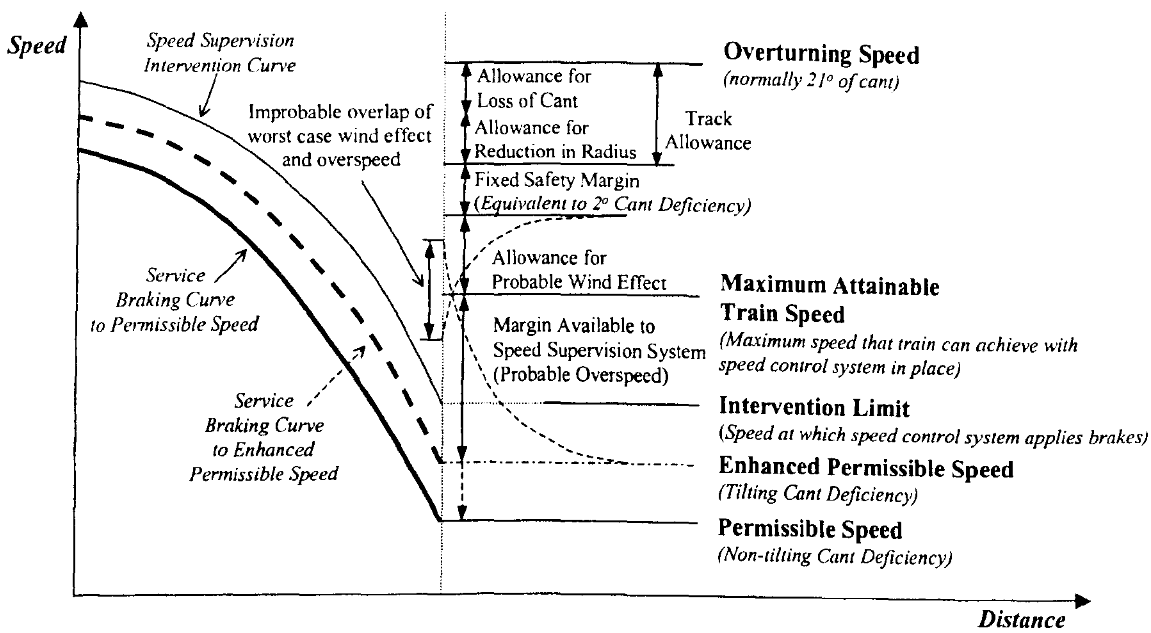


Figure 6-5: Determining the Safe Limit of Permissible and Enhanced Permissible Speeds (based on Blakeney[1] 2001, pp29-30; Cheesewright 1999, p14)

6.3.4.1 *REDUCING THE PROBABILITY OF OVERSPEED*

A number of steps can be taken to reduce the probability of overspeed under the supervision of an ATP system:

- The magnitude of any individual speed reduction can be constrained: Speed overruns become more likely when large reductions in speed are being supervised (Cheesewright 1999, p66);
- An enhanced intervention braking rate can be utilised: Overspeed probability can be reduced significantly if the brake rate on intervention is significantly better than the rate used in service braking by a driver (Cheesewright 1999, p66);
- Restrictions can be made to commence before they are actually needed: Overspeed is less likely if there is a margin between the point that the service braking curve targets for speed reduction and the beginning of the actual speed restriction (Cheesewright 1999, p66);
- The ATP system can be configured to utilise the guaranteed emergency brake as a first intervention: Where a non-vital service brake is used as the first intervention, failure of that brake application will lead to a higher overspeed level pending intervention of the guaranteed emergency brake;
- The ATP system can be configured to utilise reduced margins for warning and intervention: Lower margins would prompt the driver to apply the brake at a lower speed and reduce the maximum overspeed that can be attained following an ATP brake intervention;
- More accurate speed measurement systems can be used: Lower errors in the measurements of speed used for both ATP supervision and the drivers speedometer display would reduce the overspeed probability under worst case conditions;
- The ATP system can be configured to intervene without warning when the train exceeds a given speed as, for example, is done with the TPWS speed check: A lower intervention margin can be used if no warning margin is required, reducing the maximum overspeed that can be attained following intervention.

Reducing the magnitude of individual speed restrictions would mean, in effect, applying restrictions earlier than it is required in order to 'split' large reductions in speed into two (or more) smaller reductions. Adding in margins in this way to ensure that overspeed does not occur would reduce the efficiency of the railway's operation and is not, therefore, an ideal solution.

The use of an enhanced braking rate on intervention presents greater possibilities. According to Railway Group Standards, the choice between first intervention with the service or emergency brake is up to the train operator (Blakeney 2002, p22). This means that there is no fundamental objection or barrier to the implementation of this proposal. However, it must be done within the constraints of two major limitations: the adhesion available and the physiological effects on passengers of increased acceleration and jerk rates. The limits imposed by these constraints have already been discussed in chapter 5.

Whilst frequent use of high acceleration and jerk rates may lead to unsatisfactory perceptions of 'discomfort', this is also true of the effects of high levels of overspeeding. It is also worth noting that in practice, the worst case predictions for overspeeding would be expected to represent rare occurrences (with gradients only rarely being as high as 3% and drivers tending not to accelerate at the highest available rate as the train approaches the permissible speed). The use of higher intervention braking rates only on the rare occasions that 'worst case' conditions do occur would, therefore, seem a possibility for reducing potential overspeed in future systems.

Configuring the system to utilise the guaranteed emergency brake as a first intervention is an option provided for in the ETCS system. Where this option is adopted, the values for dV_ebi are reduced to those normally utilised for dV_sbi (Cheesewright 1999, p8). This would reduce the maximum overspeed before intervention by 0.5 km/h for a permissible speed of 0km/h and 2.5km/h for a permissible speed of 200km/h in the event of service brake failure (see Figure 6-2). If the ideal speed supervision criteria outlined in section 6.3.2 were being used, the reduction in maximum overspeed before intervention would be even more significant, at 3.3km/h. The probable levels of overspeed would, therefore, also be reduced, potentially permitting the use of enhanced permissible speeds. However, since the analysis conducted in determining the probabilities shown in Table 6-1 concluded that the overturning risk resulting from a driver failing to apply the brakes at the speed supervision system's warning margin is negligible, it must be concluded that the effect of reducing intervention limits would also be negligible. There does not, therefore, appear to be any scope in pursuing this solution further.

In order to reduce the margin provided to a driver between warning and system intervention, it would be necessary to reduce one or more of the driver's response time, the systems response time (for traction cut off and brake application) or the maximum acceleration that can occur following the warning.

In order to reduce the driver's response time, the driver can be made aware of an impending warning by indication of the maximum permitted speed in his/her cab. This is already done in the BR-ATP and ETCS systems, but not in the TASS supervision of enhanced permitted speeds. Within the design of the TASS system, it was considered that driver reaction times would already be low since the driver would be aware from his/her speedometer that the train's speed was approaching the signed EPS limit. This is also the case for ceiling speed supervision in both the BR-ATP and ETCS systems. In the light of this, it appears unlikely that any significant reduction in driver reaction time could be obtained by future systems.

System response times may be more open to improvement. Both the class 390 and class 221 have a stated traction cut off delay of 1 second. For the class 390 service and emergency brake there is an application delay of around 2 seconds, the brake then taking a further 2 seconds to build up to 90% effectiveness. This is similar to the service brake of the class 221, which takes around 2.3 seconds to apply and a further 1.3 seconds to reach 90% effectiveness. However, the class 221 emergency brake responds much more rapidly, with an application delay of only 0.2 seconds and build up delay of only 0.2 seconds to reach 90% effectiveness (Barnard[(1)] 2002, pp31 & 38). If all train brake application commencement delays could be reduced to the levels of the class 221 emergency brake (0.2s) and their brake build up delays to the minimum permitted for a comfortable 2.35m/s^3 service brake jerk rate (achieving a peak deceleration rate of, say, 0.9m/s^2 in 0.38s), supervision margins could be significantly reduced. As an example of this, if the brake application and build up delays on the class 390 were reduced to these levels, the margin required between warning and intervention for the worst case conditions (with no traction acceleration) discussed in section 6.3 would reduce by 2.2km/h to 2.4km/h. Assuming that traction acceleration could occur at maximum rate, the peak overspeed before the driver could bring the train speed under control following a warning would be reduced by the same amount to just under 7.2km/h. This would result in a reduced probable overspeed and offer the potential for increasing EPS and some TSR values.

Some of the technological solutions that can be used to reduce brake build up delays have already been discussed in section 5.6 and the feasibility of such improvements are demonstrated by the performance of the class 221 emergency brake.

The potential for reducing warning margins through the reduction of traction cut off delays is not as significant as that of reducing brake application delays. For example, a reduction in cut off delay from 1s to 0.5s would only enable reduction of the margin between warning and intervention by 0.8km/h to 8.5km/h for the worst case conditions discussed in section 6.3.1. The author has also been unable to identify any significant potential for such reductions to be made. There does not, therefore, appear to be sufficient scope to warrant pursuing this solution further.

If future ATP systems were developed with the ability to apply gradual control of the train's acceleration, it could be possible to ensure that the acceleration available to the driver reduced as the train approached the permissible speed. This would reduce the probable overspeed. If this approach were to be adopted, care would need to be taken in selecting the level of restriction that could be applied to traction performance in order to ensure that the benefits gained did not come at too significant a cost in terms of the efficient operation of the railway up to the permissible speed. If, for example, the peak acceleration rate of the train was limited to $0.22m/s^2$ when approaching a permissible speed below 70km/h, the maximum overspeed in a class 390 would reduce for the worst case conditions discussed in section 6.3 from 9.3km/h to 7km/h. If this were to be coupled with an improvement in brake application and build up delays to 0.2s and 0.38s respectively, the required margin would be further reduced to 5km/h. The author has not discovered any proposals for systems behaving in this way, but automatic control of available traction is not a totally new concept. "In London since the 1920s train acceleration, and since the 1930s train retardation, have been subject to automatic limitation of maximum values which the driver has not been able to influence. This has been done by automatically controlling the acceleration current in the case of motoring and by an inertial device in the form of a mercury switch for braking" (Maxwell 1975, p230). Intervention by control of available traction would represent a far 'softer' intervention than application of the train's brakes and the author, therefore, considers investigation of the likely capacity effects to be worthwhile. It is unlikely that significant benefits could be gained by this approach at high speeds (where train acceleration performance is already low), but significant benefits could potentially be gained at lower speeds.

A similar approach could be adopted by automatic activation of a full traction inhibit whenever the warning speed is reached or exceeded. This would not require intervention of any kind until the driver had definitely exceeded the permitted speed and the worst case conditions discussed in section 6.3, would reduce the margin required between warning and service brake intervention for a class 390 by up to 3.1km/h (assuming a maximum traction acceleration rate of $0.44m/s^2$) to only 6.2km/h.

If speed measurement errors could be reduced (or eliminated), the existing supervision criteria could be achieved with lower overspeed probabilities and without causing intervention earlier than is really necessary. This would improve the safety, efficiency (through potential for the use of higher EPS / TSR values) and passenger comfort of train movements.

Historically, speed measurement precision has typically been ±6km/h below 160km/h, according to figures quoted by Railtrack in their recommendations for systems for the supervision of enhanced permissible speeds and tilt enable (Blakeney 2001[2], p33). The performance requirements for ERTMS interoperability contain an even larger maximum figure of 6.2km/h, based on ±2km/h below 30km/h, ramping up linearly to ±12km/h at 500km/h, with a UK maximum speed of 225km/h (UNISIG 2000, p9). This requirement has been refined by Alstom Transport to permit a maximum of ±5.9km/h (based on 2km/h at speeds up to 30km/h and 2 + 2% (V-30) km/h at higher speeds) for the West Coast Main Line Train Control System (Le Vegue 2002, p7). These figures represent large (and therefore significant) potential for error. However, new trains developed for use on UK railways are required to have far smaller speedometer errors of ±2mph (3.2km/h), suggesting that far better accuracy is now possible (Barnard[1] 2002, p31). Indeed, the TASS speed measurement system (used as the basis for the calculations within this chapter) is expected to provide even greater accuracy, to 0.5km/h +0.5% of speed (that is, to within 1.7km/h at 225km/h) (Barnard[1] 2002, pp32, 38).

In the light of these developments in speed measurement accuracy, the author does not believe that further significant improvements are likely in the short term. Whilst any such developments would offer benefits to future ATP systems during high speed operation it is the author's opinion that the potential for improvement is insufficient to warrant further investigation of more accurate speed measurement techniques at the current time. However, reducing the maximum speed measurement error from 6km/h to 1.7km/h offers a direct reduction in the margin required before warning of 4.3 km/h, with a proportional reduction in probable overspeed. Therefore, the use of speed measurement systems with the improved accuracy levels of TASS is to be strongly encouraged.

It is possible to envisage the use of a service or emergency brake intervention without warning once a train has exceeded the permissible speed by a small margin (sufficient to allow for reasonable fluctuations around the permissible speed). This would remove the need for a margin between driver warning and brake intervention, reducing the probable overspeed and / or permitting the use of higher EPS / TSR values. In order to consider this proposal properly, it is necessary to understand two things. Firstly, why warnings are ever given to the driver before intervention and, secondly, how accurately a driver can reasonably be expected to follow the permissible speed without a warning.

Whether or not to provide the driver with a warning is not a question that can be resolved as a clear-cut matter of safety. It is largely determined by the operational philosophy of a given railway (based on its perception of human factors issues, particularly relating to the allocation of functions between humans and machines/automation) and the constraints of technical capability. Depending on the prevailing philosophy and technology available, there are three possible approaches that could be safely adopted:

1. If the philosophy maintains that the driver should have control of the train's movement at all times, an ATP system should ideally include accurate enough supervision criteria and means of condition assessment to ensure that a brake intervention will never be invoked if the driver is controlling the train's movement safely. Under this approach, the driver does not need to receive any indications or warnings about the ATP system's supervision criteria. The ATP can, therefore, act as a secondary safety system (or safety net) supporting the driver (as the primary safety system) in case of serious error, but otherwise remaining unseen;

2. Alternatively, if the technology required by that approach is not available, the driver must be provided with enough information about the ATP's supervision to ensure that he/she can control the train's movement not only safely, but without invoking a brake intervention where the ATP supervision is more restrictive than required. As the information displayed to the driver can mislead in the event of ATP failure or error, the ATP system becomes the primary safety system for the supervised actives, with the driver in a secondary role. This means that the integrity of the system must be higher;
3. A third approach becomes possible if the philosophy permits control of the train to be removed from the driver and taken on by the ATP system where supervision criteria are exceeded, even if the limits of safety have not been. Under this approach, the driver does not need to receive any indications or warnings about the ATP system's supervision criteria. The system, therefore, can act as a secondary safety system (or safety net) supporting the driver (as the primary safety system) in case of deviation from supervised behaviour, but otherwise remaining unseen.

With the first of these approaches, if the ATP is to ensure both safety and efficiency of operation, it must be able to assess and react to conditions such as adhesion and train performance in the same way as an experienced driver. Unfortunately, an ATP that can guarantee meeting this performance requirement has yet to be developed. Therefore, this approach is not currently feasible. In the opinion of the author, this is likely to remain the case for some time, pending the development of far superior technology to that available today.

With the second approach, the driver will require target speed indications and/or warnings to be provided with sufficient margin to allow brake intervention to be avoided by prompt reaction. If the margins are not sufficient for this, they become superfluous (the warning merely acts as a notification that brake intervention is about to happen). Unfortunately, this approach can lead to the driver becoming dependent on, or at least influenced by, the indications and warnings given. If the philosophy of making the driver the primary control system is to be maintained, the information displayed to the driver must be kept to a minimum and not distract the driver from performing his/her normal role. It is this approach that has been adopted in the UK for BR-ATP, and ETCS (with target speed information) and TASS (with no target speed information).

With the third approach, it is accepted that the ATP system may, on occasions, be more restrictive than the actual conditions (such as adhesion and train performance) permit. When this occurs, the driver is expected to drive in accordance with pre-defined behaviour, rather than to his/her interpretation of the limits of safe operation. If the driver allows the train to deviate from this expected behaviour, control of the train's movement will be taken over by the ATP, even if the driver has not actually done anything unsafe. This may require revised behaviour by drivers (particularly on days where conditions would permit better performance than is usual). A main disadvantage of this approach is that it may not be possible to make full use of the train's potential safe performance where that exceeds the supervised performance. Another is the potential for loss of respect for the system by drivers if it intervenes unnecessarily on a regular basis. This general approach has been adopted in the UK for the TPWS system, with some evidence of revised driver behaviour being required (particularly on approach to terminal stations and when operating in conjunction with BR-ATP).

Whilst the technology required for the first approach is not currently available, it is possible to envisage an ATP system operating with the third approach that would be capable of far more intelligent supervision than TPWS and could achieve supervision close to the ideal of the first approach. As technology develops, it should become possible to draw closer and closer to the ideal, minimising the number of unnecessary interventions and minimising the constraints that need to be placed on safe driver behaviour. To the author, this appears to be an approach well worth further consideration.

The overturning risk resulting from a driver failing to apply the brakes at a speed supervision system's warning margin has been shown to be negligible. If probable overspeed is to be reduced, rather than increased, any intervention without warning would, therefore, need to occur within the normal allowance for driver's reaction following a warning. In accordance with the ideal speed supervision criteria outlined in section 6.3.2[4], the intervention would need to be within 7.9km/h of the permitted speed. It has already been noted in section 6.3 that a mechanical speedometer typically requires a tolerance for driver control of 4km/h. Assuming that any other type of speedometer would require no greater tolerance, this can be taken as the absolute minimum margin required between permissible speed and intervention without a warning. Hence, a reduction in probable overspeed or increase in EPS / TSR of up to 3.9km/h could be possible by not including a warning before intervention. However, it is the opinion of the author that such a development would be unacceptable to the operation of the railway unless the intervention speed were to be set above that usually required for a warning to be given, rather than above the minimum tolerance for driver control. This reduces the potential for reduction in probable overspeed to 2.1km/h for the equivalent of a 5.8km/h warning margin. In conjunction with some of the other proposals made in this section, a reduction in probable overspeed by this amount would be worth further consideration.

6.4 SUMMARY

For conventional trains, the predominant factor in determining permanent speed restrictions is passenger comfort. This means that large margins are available between the defined permitted speed and speeds at which overturning and / or derailment become serious concerns. Since ATP systems do nothing to address passenger comfort directly at a given speed, there would not seem to be a case for allowing signed permitted speeds to be increased for ATP fitted trains in general (to do so must be expected to cause passengers discomfort). However, there would seem to be a good case for fine tuning ATP supervision criteria in order to reduce probable overspeed. This could potentially support the raising of some temporary speed restrictions and / or enhanced permissible speeds for tilting trains, where the predominant factor in defining the acceptable speed limits is the risk of overturning or derailment.

In accordance with this objective, a number of areas that offer scope for future improvements in ATP supervision criteria have been identified within this chapter. In particular:

- The use of higher braking rates on ATP intervention;
- The gradual control of traction acceleration by limiting the response to a driver's demand (particularly at low speeds) as the train approaches the permissible speed, or the complete cutting of traction following an overspeed warning;
- The use of speed measurement systems with improved accuracy, as proposed for the TASS system;

[4] Section 6.3.2 suggested that a driver should be allowed two seconds reaction time, with no traction acceleration, following a warning at 5.8km/h overspeed

- Elimination of warnings prior to brake intervention, thus enabling reduced margins between permitted and intervention speeds.

It is unfortunate, that the current proposals for ATP supervision margins on ETCS and those implemented for TASS are already smaller than appears necessary to permit reasonable levels of flexibility in the driving of fitted trains at low speeds. In the author's opinion, testing prior to ETCS final implementation will lead to changes being made to the proposed supervision criteria in consequence of this. For future systems, however, the technology does exist to enhance train performance and ATP supervision sufficiently to overcome the problems anticipated for TASS and ETCS. Since the margins already appear to be larger than required at high speeds, such developments could lead to significant reductions in probable overspeed at all permitted speeds.

7 THEORETICAL CONTROL SYSTEM CAPACITY

7.1 INTRODUCTION

A basic consideration of the theoretical capacity of railway operation was commenced in chapters 3 and 4. This looked at 'train following headway' for fixed and moving block signalling systems in accordance with traditional UK main line signalling practice, utilising equi-spaced signals with nominal length overlaps. However, railway operation cannot be sustained with all trains travelling at line speed as assumed in train following headway. Trains must stop to allow passengers to board and alight at stations. This requires a departure from the line speed during braking and subsequent acceleration, as well as a dwell of some duration in the station. Trains may also be required to operate at speeds other than the line speed for other reasons, such as speed restrictions or the limitations of the train's performance. Therefore, it may be more useful to consider 'point headways' that account for variations in train speed.

In addition to this, not all signalling installations are in accordance with traditional UK main line signalling practice. Signalling on LUL is traditionally not equi-spaced and has overlaps long enough to stop a train from its maximum achievable speed. Other signalling applications utilise equi-spaced signals with overlaps made up of a full signal section. The adoption of these different approaches affects the headway that can be achieved by a railway.

In this chapter the author builds on the train following headway equations derived in chapters 3 and 4, developing equivalent equations for point headways representative of the range of signalling systems currently in use in the UK:

- Equi-spaced signals with nominal length overlaps;
- Optimally spaced signals, in accordance with traditional LUL practice;
- Equi-spaced signals with a full block section overlap.

7.2 MULTIPLE ASPECT TRAIN FOLLOWING HEADWAYS

The detailed consideration of train following headways given in chapter 3 concluded that for two aspect signalling the headway time is given by:

$$Ht_2 = \frac{B+X+S+O+L}{V} \qquad \text{Equation 3-10}$$

In chapter 4 it was shown that, for three or more aspects, the headway time is given by:

$$Ht_n = \frac{\left[\frac{(n-1)B}{n-2}\right]+S+O+L}{V} \qquad \text{(where n>2)} \qquad \text{Equation 4-1}$$

As the number of aspects is increased, the ultimate 'infinite' aspect system is equivalent to operation using moving block. The headway time for moving block was found in chapter 4 to be given by:

$$Ht_\infty = \frac{B+S+O+L}{V} \qquad \text{Equation 4-2}$$

Where O is a safety margin and S an allowance for driver reaction to in-cab signalling changes.

These equations only represent the headway that can be achieved by operation of the railway at the designed line speed. However, trains may actually travel safely over a section of fixed block signalling at any speed up to the designed line speed. It can, therefore, be useful to consider the effects on headway of lower speed operation.

7.2.1 TRAIN FOLLOWING BELOW DESIGNED LINE SPEED:

As trains travel at speeds lower than the designed line speed, the achieved headway time will increase. This can be demonstrated by developing the equations for headway distance in section 3.11.5, as follows:

$$H_n = (n-1)d + S + O + L \qquad \text{(for n>2, where d = B/(n-2))} \qquad \text{Equation 3-7}$$

Since B is the braking distance from the line speed:

$$B = \frac{V^2}{2b} \qquad \text{(where b = braking deceleration rate; V = line speed)} \qquad \text{Equation 7-1}$$

The equation can be developed to give the minimum 'n' aspect headway distance:

$$H_n = \frac{V^2(n-1)}{2b(n-2)} + S + O + L \qquad \text{(for n>2)} \qquad \text{Equation 7-2}$$

When travelling at an actual speed given by V_{act}, this can be converted into a headway time:

$$H_{tn} = \frac{V^2(n-1)}{2bV_{act}(n-2)} + \frac{S+O+L}{V_{act}} \qquad \text{(for n>2)} \qquad \text{Equation 7-3}$$

For 4 or more aspect signalling, if an actual train's speed is sufficiently below the design speed, it may become possible to stop safely within a reduced number of signal separations. This is typically the case for local passenger services operating over a line designed for high-speed passenger trains. It may also be the case when the service is recovering from a serious perturbation. When this occurs, trains can proceed past cautionary aspects without the driver needing to adjust the train's speed (Anon. 1967, p607; Nock 1980, p11; Pope 1975, p16). Since under these conditions the train in front is not impeding the following train, headway is still being achieved. The equation for headway time can, therefore, be expanded as follows:

3 Aspect: $$H_{t3} = \frac{V^2}{bV_{act}} + \frac{S+O+L}{V_{act}} \qquad \text{Equation 7-4}$$

4 Aspect: $$H_{t4} = \frac{3V^2}{4bV_{act}} + \frac{S+O+L}{V_{act}} \text{ for } V_{act} > \frac{V}{\sqrt{2}} \qquad \text{Equation 7-5}$$

$$H_{t4(3)} = \frac{V^2}{2bV_{act}} + \frac{S+O+L}{V_{act}} \text{ for } V_{act} \le \frac{V}{\sqrt{2}} \qquad \text{Equation 7-6}$$

5 Aspect: $$H_{t5} = \frac{2V^2}{3bV_{act}} + \frac{S+O+L}{V_{act}} \text{ for } V_{act} > V\sqrt{\frac{2}{3}} \qquad \text{Equation 7-7}$$

$$H_{t5(4)} = \frac{V^2}{2bV_{act}} + \frac{S+O+L}{V_{act}} \text{ for } \frac{V}{\sqrt{3}} \le V_{act} \le V\sqrt{\frac{2}{3}} \qquad \text{Equation 7-8}$$

$$H_{t5(3)} = \frac{V^2}{3bV_{act}} + \frac{S+O+L}{V_{act}} \text{ for } V_{act} \le \frac{V}{\sqrt{3}} \qquad \text{Equation 7-9}$$

Similar equations can be derived for any number of aspects, and ultimately, for moving block operation (when 'n' tends to ∞):

$$H_{tm} = \frac{V^2}{2bV_{act}} + \frac{S+O+L}{V_{act}}$$ Equation 7-10

Since moving block headway is not constrained by pre-designed signalling equipment locations, it is a dynamic feature of the train's current speed. This means that V_{act} is always equal to V and the moving block headway at any speed is, therefore, always given by equation 4-2.

The equations derived in this section can be plotted to give the train following headway curves shown in Figure 7-1:

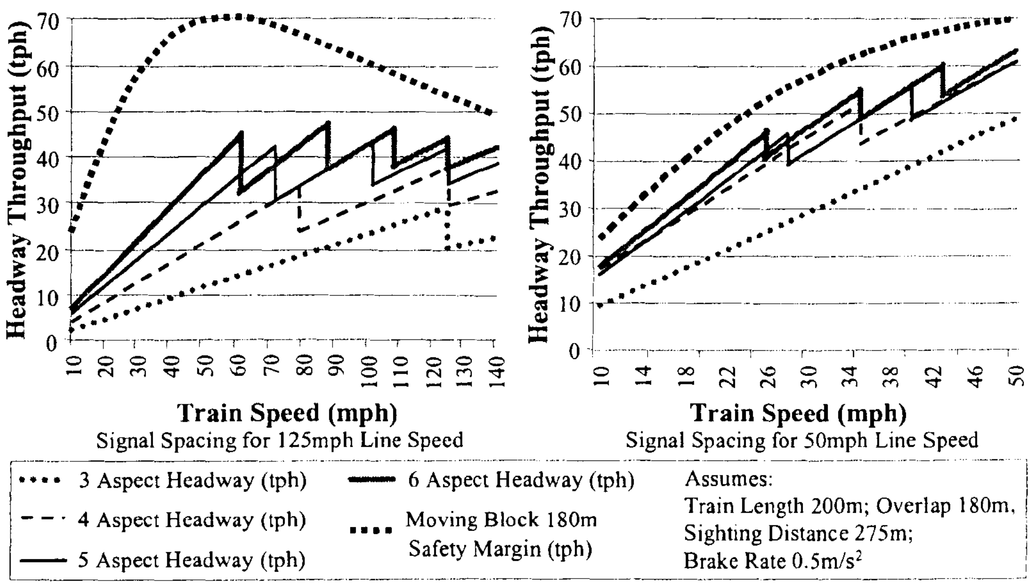


Figure 7-1: Variation in Train Following Headway with Speed, for Fixed Design Speed

Several interesting performance characteristics can be observed in the results of Figure 7-1:

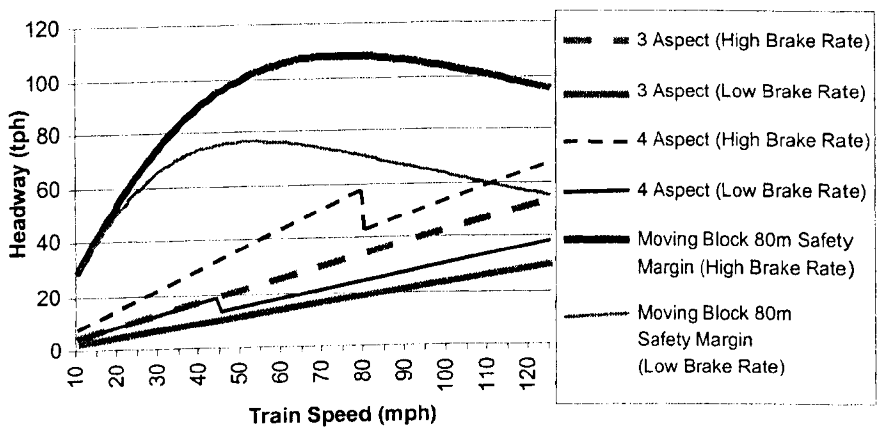
- For fixed block systems, a significant step improvement in headway can be achieved when low speeds of operation permit the less restrictive aspect(s) to be ignored by drivers;
- When comparing the headways achieved by fixed block systems with differing numbers of aspects, the system with the lower number of aspects may actually offer the better headway at some speeds of operation. For example, a five aspect system designed for 125mph operation but actually operating at 70mph will provide safe train separations if the top two aspects are ignored by the driver (i.e. the driver can treat it as a three aspect system). For the same design and actual speeds a six aspect system will also be able to provide safe train separations if the top two aspects are ignored by the driver (i.e. the driver can treat it as a four aspect system). For a very narrow speed band, the five aspect system therefore offers the better headway;
- When the control system uses a high number of aspects or moving block for designed line speeds above the optimal throughput speed for that number of aspects, the achievable headway may actually increase when slower operation permits drivers to ignore the top aspects(s). For example, a six-aspect system with a designed line speed of 125mph will achieve better headways in the region of 90mph (when the driver can treat it as a 4 aspect system) than it will at 125mph. This effect is far clearer for moving block, which achieves optimal performance in the region of 60mph regardless of the designed line speed;

- As a moving block system allows continuous optimisation for operation at any speed, the operation of the railway under moving block always offers improved performance at reduced speeds when compared to fixed block multiple aspect systems. This improvement in performance is more noticeable at higher designed line speeds.

Where vehicles operating on the line have different braking performance as well as different operating speeds, the designed signal spacing may impose restrictions on the acceptable operating speeds of the trains with poorer braking performance. If, for example, a section of fixed block signalling is designed for high-speed (125mph) passenger services but is also used to operate slow freight trains with poor braking performance, the maximum speed of the freight trains must be limited to that from which they can safely stop within the allowed braking distance. The headway that would result from operating the line purely with poor braking performance freight trains can be seen in Figure 7-2, compared with that achieved by operating purely high speed services with higher brake performance.

The headway that would be achieved in practice on such a section of line would dependent on service patterns for the various trains in use. Further discussion of the effects of service patterns can be found in section 2.5 and will not be repeated here. It is sufficient to note that, as represented in Figure 7-2:

- The introduction of stock with significantly poorer braking performance will markedly reduce the throughput of trains that can be achieved on a section of line;
- This effect is far less pronounced where train separation is achieved by use of a moving block, rather than a fixed block signalling system.



Train Length 200m; Overlap 180m, Sighting Distance 275m;
High Brake rate 1.0m/ s^2, Low Brake Rate 0.5m/s^2
Signal separations designed for trains with high brake rate, operating at 125mph

Figure 7-2: Effect of Brake Rate and Speed on Train Following Headway

7.3 POINT HEADWAYS

So far, the equations derived for signalling system headways have all considered the following headway of trains travelling at a constant speed. However, the operation of any railway requires trains to slow down from line speed for station stops. It may also require trains to travel through diverging or converging junctions. The headway of trains undergoing such deviations from plain line operation at line speed (the point headway) can also be analysed in the same way.

There are several cases to consider for point headways: the case where consecutive trains stop; the case where non-stopping trains follow stopping trains; the case where stopping trains follow non-stopping trains; converging junctions; diverging junctions (with a variety of high and low speed approach conditions).

7.3.1 ALL TRAINS STOPPING

If all trains stop at a station, the headway time is determined by the worst-case headway of all signals in the station area. If the train travels at the permitted speed throughout a signal's route, the headway time for that signal will be equal to that of plain line. If it travels below the permitted speed at any point within a signal's route, the headway time of that signal will be greater than that of plain line. This means that all signals from the first 'home' (station protection) signal up to and including the first signal in the route from which the train regains the permitted speed should, technically, be considered. In practice, however, trains approaching a station stop do not require green aspects to proceed unaffected by the train in front (see Figure 7-3). Whilst trains' acceleration rates are generally lower than their braking rate, the dwell in the station (which may be from 20 seconds to a few minutes; Barter 2001, p9) means that the signals controlling the exit from a station area have lower headways than those controlling the arrival. The delay that occurs as a train enters and stops at a station also gives time for the preceding train to clear the area. In consequence of this, the first home signal is the only signal that actually needs to be analysed (that is, three signals before the station starter in 4 aspect territory and two signals before it in 3 aspect territory). A proof of this can be found in Appendix D.

Generally the headway time will be made up of four components:

Cruising time + deceleration time + dwell time + acceleration time Equation 7-11

If we begin by considering a three aspect example, the headway distance for signal 1 (the first home signal) will remain as already derived in section 3.11.3 for train following headway distance (see Figure 7-3).

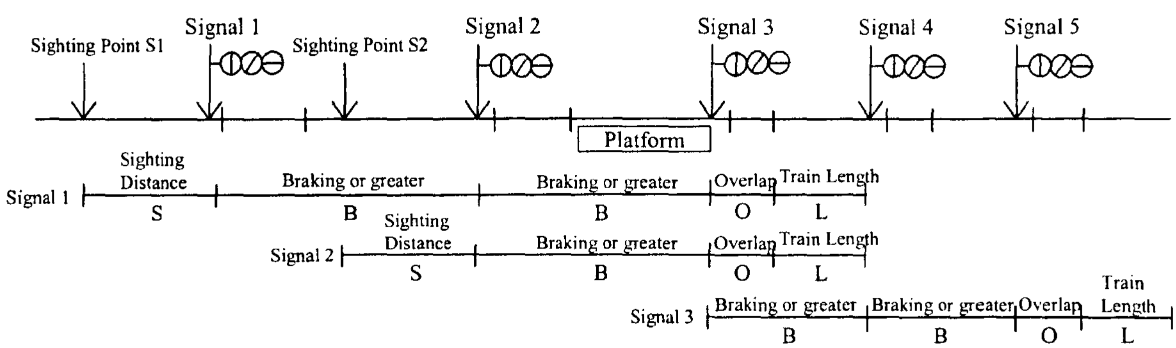


Figure 7-3: 3 Aspect Signal Headways (Station Stop)

However, in order to convert this to a headway time, the variations in the speed of the train during its passage through the route must be considered. During the sighting distance and first signal separation the train can travel at line speed. It must then travel a signal separation (braking distance) in the process of braking, in order to stop at signal 3. After a dwell time, the train must then accelerate out of the platform and travel far enough to clear the overlap of signal 3 before the next train arrives at the sighting point of signal 1.

$$H_{t3(Stopping\ S1)} = \frac{S+B}{V} + BrakingTime + Dwell + TimeToTravelClearofS3Overlap \quad \text{Equation 7-12}$$

Whilst accelerating to line speed, the train travels a distance $V^2/2a$ in time V/b. Calculating the time to travel clear of Signal 3 overlap is more complex, since it has two conditions: where the train reaches line speed before clearing the overlap and where it does not. Whilst accelerating to line speed, the train travels a distance $V^2/2a$ in time V/a. The distance from the station stop location to the end of the overlap is given by O + L. The time to clear the overlap is, therefore:

$$t = \sqrt{\frac{2S}{a}} = \sqrt{\frac{2(O+L)}{a}} \quad \text{if line speed is not reached} \qquad \text{Equation 7-13}$$

else:

$$t = \frac{V}{a} + \frac{O+L-\frac{V^2}{2a}}{V} = \frac{V}{2a} + \frac{O+L}{V} \quad \text{if line speed is reached} \qquad \text{Equation 7-14}$$

Where line speed is reached only if:

$$V \leq \sqrt{2a(O+L)} \qquad \text{Equation 7-15}$$

This gives the headway time to be:

$$H_{t3(Stopping\ S1)} = \frac{S}{V} + \frac{3V}{2b} + Dwell + \begin{cases} \sqrt{\frac{2(O+L)}{a}} & \text{For } V \geq \sqrt{2a(O+L)} \\ \frac{V}{2a} + \frac{O+L}{V} & \text{For } V \leq \sqrt{2a(O+L)} \end{cases}$$

Equation 7-16

If the case of four aspect signalling is considered, it can be seen that the headways for signals 1, 2, 3 and 4 will again all be different (as shown in Figure 7-4).

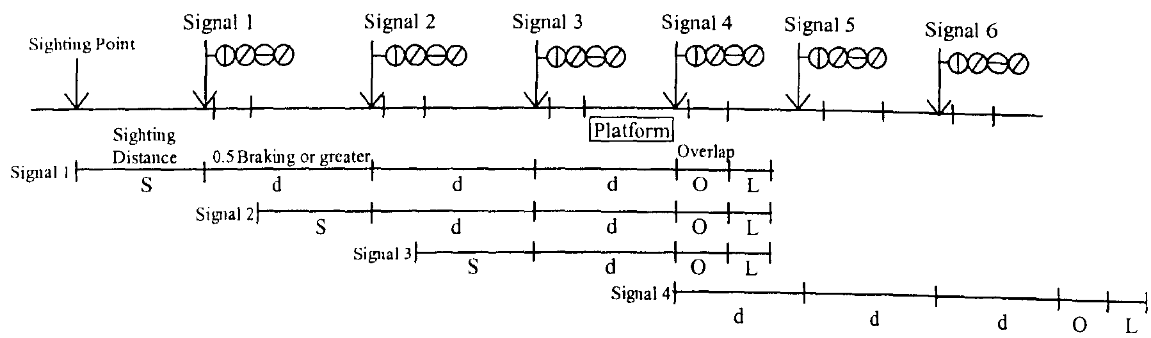


Figure 7-4: 4 Aspect Signal Headways (Station Stop)

Once again, the critical headway time will be that for signal 1(the first home signal):

$$H_{t4(Stopping\ S1)} = \frac{S+d}{V} + Braking\ Time + Dwell + Time\ To\ Travel\ Clear\ of\ S4\ Overlap \qquad \text{Equation 7-17}$$

Since d is equal to ½B, or $V^2/4b$, this gives a headway time of:

$$H_{t4(Stopping\ S1)} = \frac{S}{V} + \frac{5V}{4b} + Dwell + \begin{cases} \sqrt{\frac{2(O+L)}{a}} & \text{For } V \geq \sqrt{2a(O+L)} \\ \frac{V}{2a} + \frac{O+L}{V} & \text{For } V \leq \sqrt{2a(O+L)} \end{cases}$$

Equation 7-18

In more general terms, the limiting headway time for an 'n' aspect signalling system is given by:

$$H_{tn(Stopping)} = \frac{S}{V} + \frac{V(2n-3)}{b(2n-4)} + Dwell + \begin{cases} \sqrt{\frac{2(O+L)}{a}} & \text{For } V \geq \sqrt{2a(O+L)} \\ \frac{V}{2a} + \frac{O+L}{V} & \text{For } V \leq \sqrt{2a(O+L)} \end{cases} \quad \text{(for n > 2)}$$

Equation 7-19

This is always greater than the plain line headway time by a factor:

$$
= \frac{V}{2b} + Dwell + \begin{cases} \left\{ \sqrt{\frac{2(O+L)}{a}} - \frac{O+L}{V} \right\} & \text{For } V \geq \sqrt{2a(O+L)} \\ \left\{ \frac{V}{2a} \right\} & \text{For } V \leq \sqrt{2a(O+L)} \end{cases}
$$

Equation 7-20

For a moving block system, which is the equivalent to a fixed block system with an infinite number of aspects, the limiting headway time becomes:

$$
H_{t\,m\,(Stopping)} = \frac{S}{V} + \frac{V}{b} + Dwell + \begin{cases} \left\{ \sqrt{\frac{2(O+L)}{a}} \right\} & \text{For } V \geq \sqrt{2a(O+L)} \\ \left\{ \frac{V}{2a} + \frac{O+L}{V} \right\} & \text{For } V \leq \sqrt{2a(O+L)} \end{cases}
$$

Equation 7-21

The relationship between achievable point headway and line speed when all trains stop can be seen in Figure 7-5, for a typical sighting time, braking and acceleration rates.

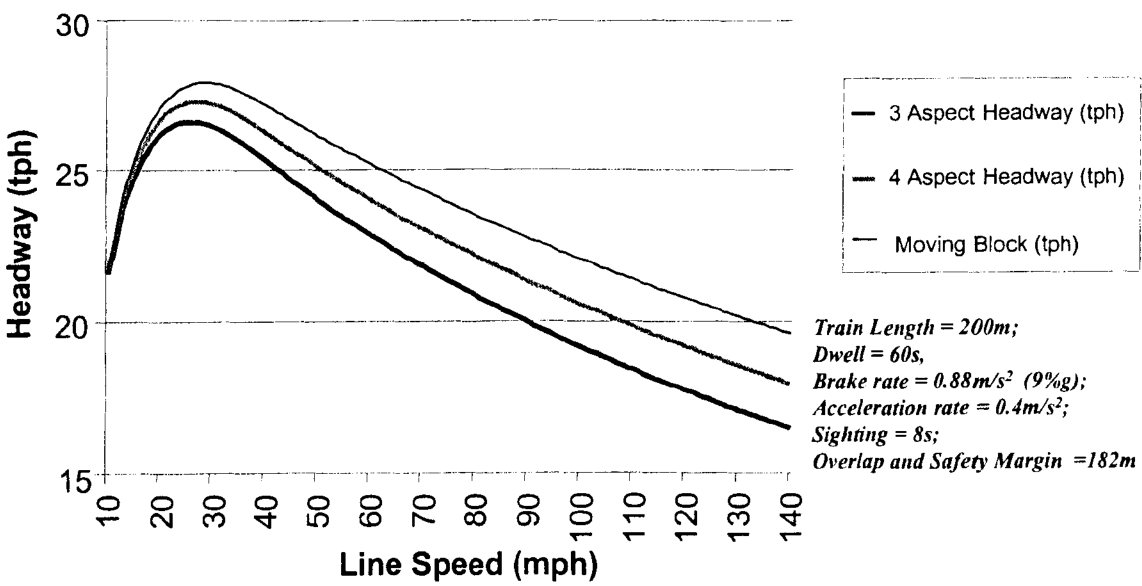


Figure 7-5: Point Headways, All Trains Stopping

7.3.2 STOPPING / NON-STOPPING

If the lead train stops at a station and the following train passes through without slowing down, the headway is determined by the run times of the two trains. This is represented in Figure 7-6.

The critical headway time represented in Figure 7-6 is the time that the following train must be behind the lead train at the start of the scenario in order to ensure that it is at least plain line headway behind the front of the lead train at the end of the scenario. It is, therefore, given by the sum of the plain line headway time and stopping train run time, less the non-stopping train run time. The start of the run times is the location at which the stopping train begins to slow from the line speed. The end of the run times is the location at which it regains the line speed following the station stop.

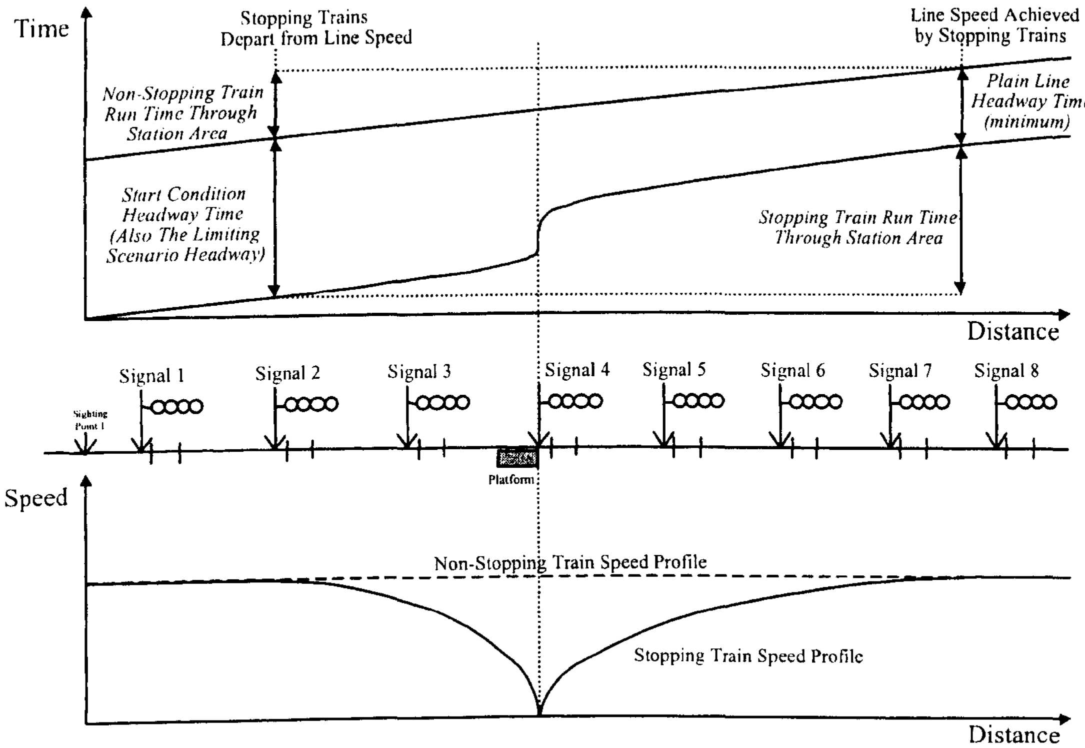


Figure 7-6: Headway for Stopping followed by Non-Stopping Train

Depending on the line speed and train dynamics, both the start and end locations will vary. The only point in the scenario that is fixed for all conditions is the station stop location. It is necessary, therefore, to work back from the station stop location in order to determine the start location and forward from it to work out the end location. This can be done by:

- Determining the plain line headway time:

$$Ht_n = \frac{(n-1)d + S + O + L}{V} \qquad \text{Equation 3-11}$$

$$\text{where } d = \frac{V^2}{2b(n-2)} \text{ and } n>2 \qquad \text{Equation 7-22}$$

- Determining the stopping distance and time taken from the permitted speed to rest:

$$Stopping\,Dist = \frac{V^2}{2b} \text{ and } Stopping\ time = \frac{V}{b} \qquad \text{Equation 7-23}$$

The start location will be this distance before the station stop;

- Determining the acceleration distance and time taken from rest to the permitted speed:

$$Acceleration\,Dist = \frac{V^2}{2a} \text{ and } Acceleration\ time = \frac{V}{a} \qquad \text{Equation 7-24}$$

The end location will be this distance after the station stop;

- The stopping train run time will be equal to the sum of the stopping time, acceleration time and any station dwell time:

$$Stopping\,Run\,Time = \frac{V}{b} + \frac{V}{a} + Dwell \qquad \text{Equation 7-25}$$

- The non-stopping train run time will be the time taken to travel a distance equal to the sum of the stopping distance and acceleration distance, when travelling at the permitted speed:

$$Non-Stopping\ Run\ Time = \frac{V}{2b} + \frac{V}{2a} \qquad \text{Equation 7-26}$$

Therefore, the limiting headway time for a stopping train to be followed by a non-stopping train is given by:

$$Ht_{n(Stop/Non-Stop)} = \frac{V(2n-3)}{2b(n-2)} + \frac{S+O+L}{V} + \frac{V}{2a} + Dwell \ \text{(where n>2)} \qquad \text{Equation 7-27}$$

This is always greater than the plain line headway time by a factor:

$$= \frac{V}{2b} + \frac{V}{2a} + Dwell \qquad \text{Equation 7-28}$$

Where $V \leq \sqrt{2a(O+L)}$, the headway time of a stopping train followed by a non-stopping train, is the same as that of consecutive stopping trains. Where V is larger than this, the headway is longer than that of consecutive stopping trains.

7.3.3 NON-STOPPING / STOPPING

If the lead train passes straight through without slowing down at a station and the following train stops, the headway is again determined by the run times of the two trains. This is represented in Figure 7-7.

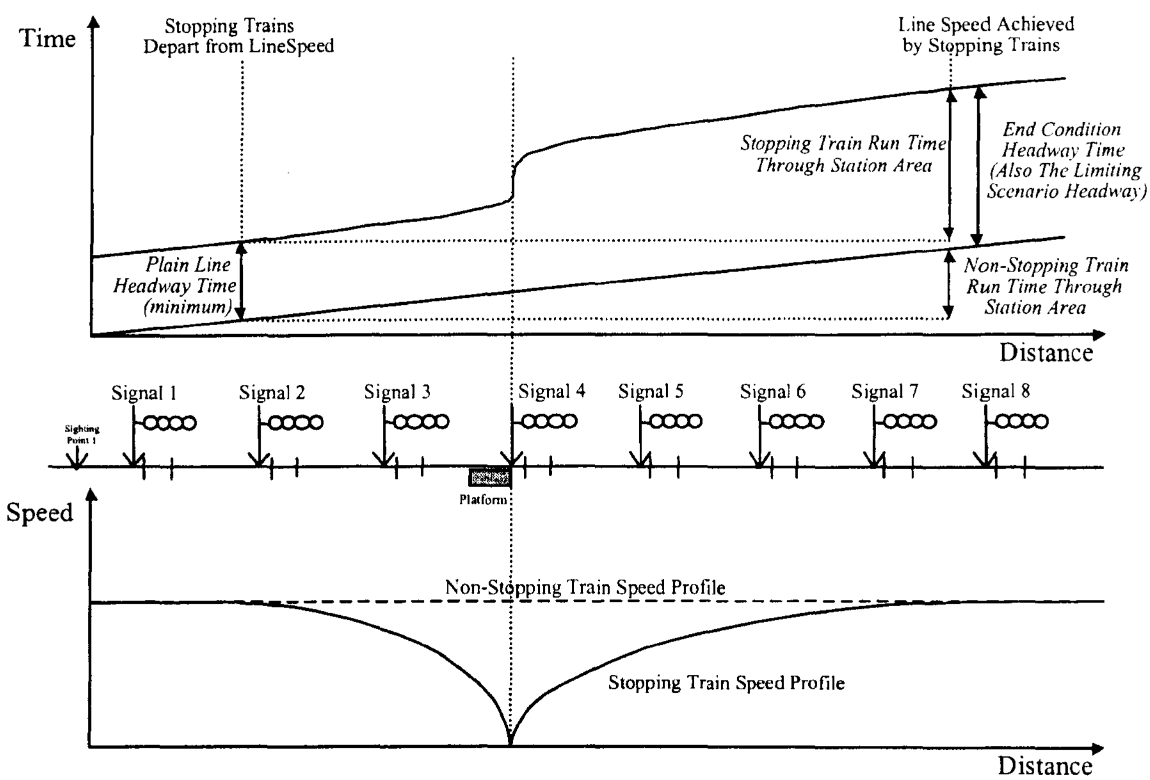


Figure 7-7: Headway for Non Stopping followed by Stopping Train

The critical headway time represented in Figure 7-7 is the time that the following train would be behind the lead train at the end of the scenario if it started plain line headway behind the front of the lead train at the start of the scenario. It is, therefore, again given by the sum of the plain line headway time and stopping train run time, less the non-stopping train run time. This means that the headway time is identical to that for a non-stopping train following a stopping train:

$$Ht_{n(Non\text{-}Stop/Stop)} = \frac{V(2n-3)}{2b(n-2)} + \frac{S+O+L}{V} + \frac{V}{2a} + Dwell \qquad \text{(where n>2)} \qquad \text{Equation 7-29}$$

This is always greater than the plain line headway time by a factor:

$$= \frac{V}{2b} + \frac{V}{2a} + Dwell \qquad \text{Equation 7-30}$$

Once again, where $V \leq \sqrt{2a(O+L)}$, the headway time of a stopping train followed by a non-stopping train is the same as that of consecutive stopping trains. Where V is larger than this, the headway time is longer than that of consecutive stopping trains.

7.3.4 STATIONS WITH PLATFORMS NOT AT A SIGNAL

All of the scenarios outlined so far for station stops have assumed that a signal will be located at the proceed end of all station platforms. In practice, whilst this is often the case, it may not always be so. If the platform were to be located in mid section, this would have a material impact on individual signal headways through the station area.

The speed / distance run profile of trains through the station area would retain the same basic shape – the only difference being the point at which the braking curve commences. In consequence of this, the headway for the scenarios involving one stopping and one following train would remain the same. In the case of consecutive stopping trains, the first home signal would also remain the critical signal for headway purposes. However, the headway time of this signal would be expected to increase, in recognition of the fact that average train speeds through the signal's route would be lower, the distance from station stop to clearing the starter overlap being longer (Pope 1975, p25). This is represented in Figure 7-8.

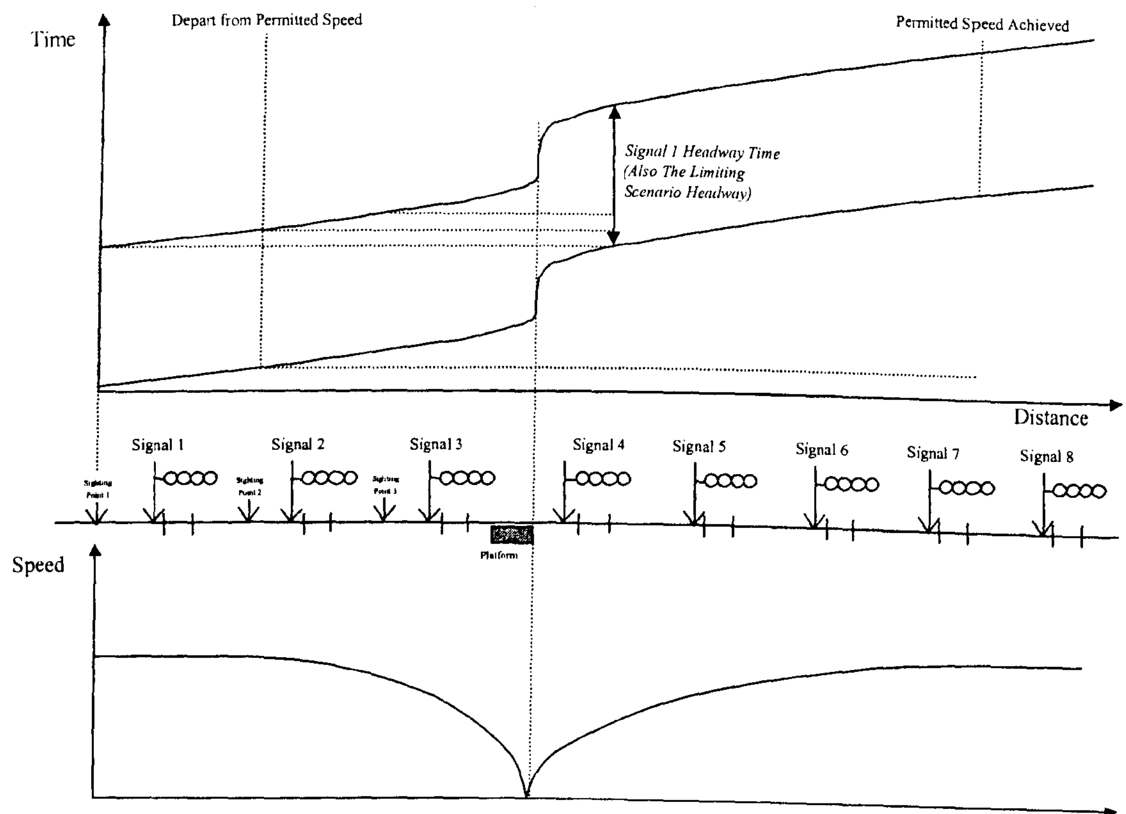


Figure 7-8: Headway for Consecutive Stopping Trains with Mid-Section Platform

For simplicity, the case of station starters being located at the end of platforms will be the only one expanded in detail within this chapter. However, the same principles used to describe headways for a station stop at a signal could be adapted to describe a stop at any location.

7.3.5 SPEED RESTRICTIONS

The headway impact of speed restrictions is again dependent on signal locations and the performance characteristics of the train being considered. When signal spacing is determined, a number of factors must be considered:

- The spacing must always be sufficient to ensure that all trains using the route can brake to rest between the first cautionary aspect encountered and a red signal;
- The spacing must not exceed the maximum permitted spacing (see section 8.2);
- The spacing must always support the headway required through that section of line.

In accordance with these principles, the separation between signals should, ideally, be based on the braking distance to rest from the maximum speed permitted (the PSR) on approach to each signal. However, where the impact on achievable headway of transition between speed restrictions is too great, the braking distance may be based on maximum attainable speeds. On Network Rail infrastructure, the use of attainable speeds in determining speed restrictions is by exception only, due to the limitations that then become imposed on train performance – any increase in acceleration performance requiring re-design of the signalling (Fleming 2000, p6). However, on LUL (which operates with homogeneous fleets of trains and typically replaces both signalling and rolling stock during the upgrade of any line), it is the normal practice.

7.3.5.1 RELAXING SPEED RESTRICTIONS – FIXED BLOCK SIGNALLING

If trains approach a PSR area from a lower speed restriction, the increased signal separation would be expected to apply for all signals within the restricted area. Ideally, this would mean that all signal separations within the area are sufficient for braking from the higher permitted speed. This is represented in Figure 7-9 for speeds at which the plain line headway of the lower restriction is better than that of the higher restriction. The constraining headway in this example can be seen to be that of the first signal within the restricted area. This signal has the increased braking distance required by the higher speed of operation, but the train travels through a part of its route at lower speeds. It, therefore, has a longer headway time than the signals with routes further into the area of the increased restriction.

As has already been discussed in chapter 3, the relationship between speed and braking distance means that, at some speeds, the headway of the higher speed will be longer than that of the lower speed. In such cases, as represented in Figure 7-10, the constraining headway can still be seen to be that of the first signal within the restricted area. This signal again has the increased braking distance required by the higher speed of operation, but the train travels through a part of its route at lower speeds.

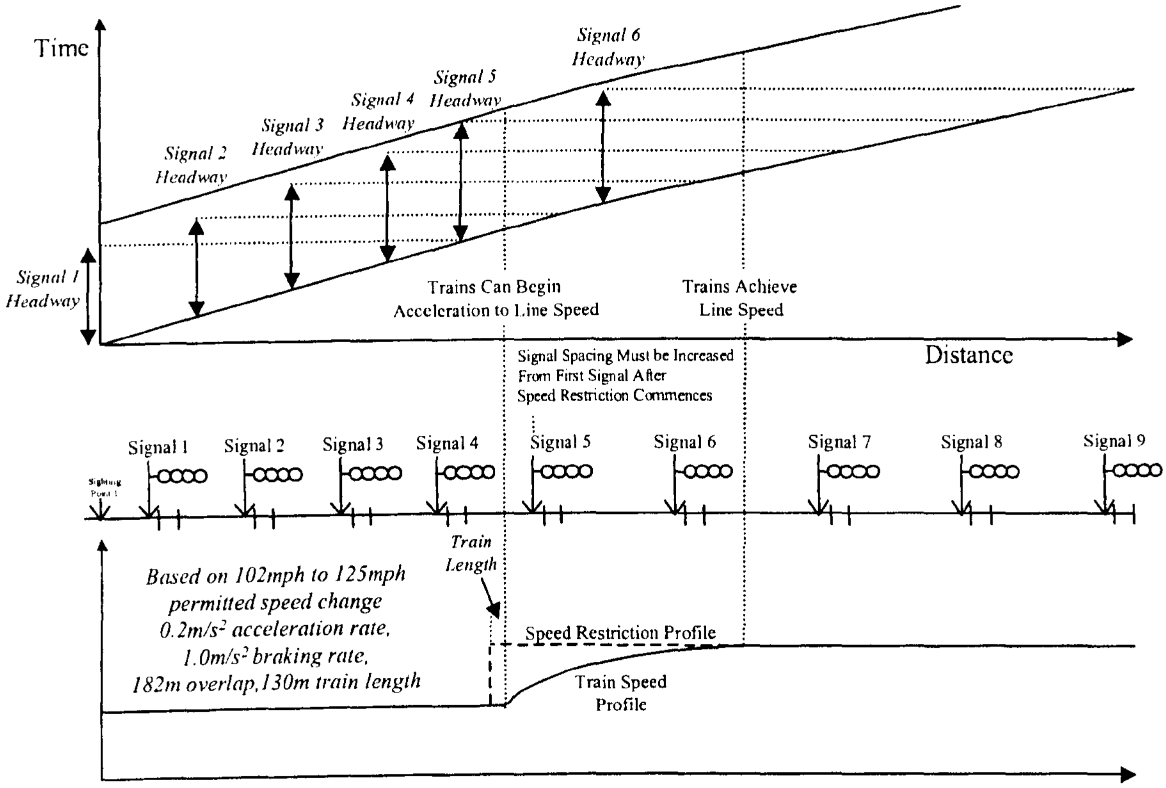


Figure 7-9: Headway for Increasing Speed Restriction at High Speeds

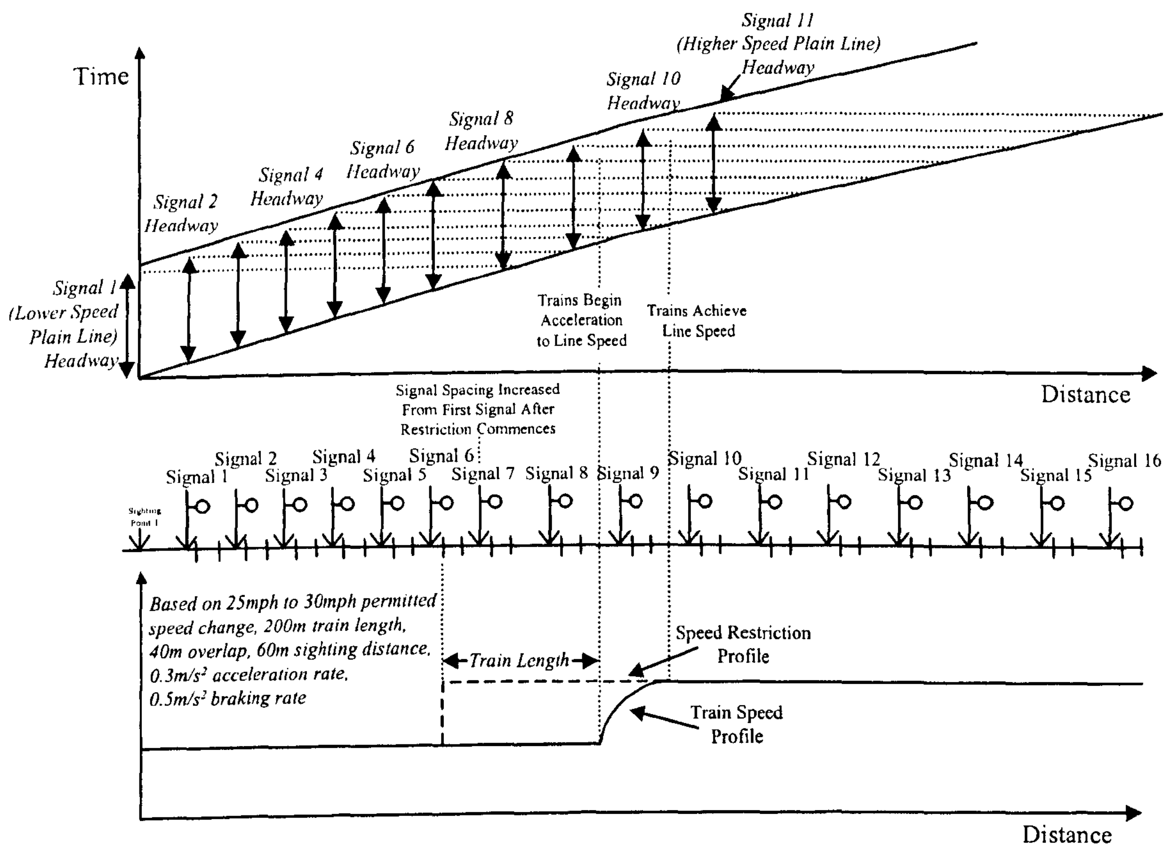


Figure 7-10: Headway for Increasing Speed Restriction at Low Speeds

In practice, trains will not be able to increase their operating speeds instantaneously to a higher permitted speed on leaving an area with a lower restriction. If the best acceleration rate of any train operating on the line is considered, maximum attainable speeds following entry to the area of the higher speed restriction can be determined and used to locate signals more optimally (closer together) just within the area, whilst still providing sufficient braking distance, reducing the impact on headway. The result of doing this is to make the minimum headway time for an increase in speed restriction equal to the longer of the plain line train following headway times for operation at the higher and lower speed restriction values. This is represented in Figure 7-11.

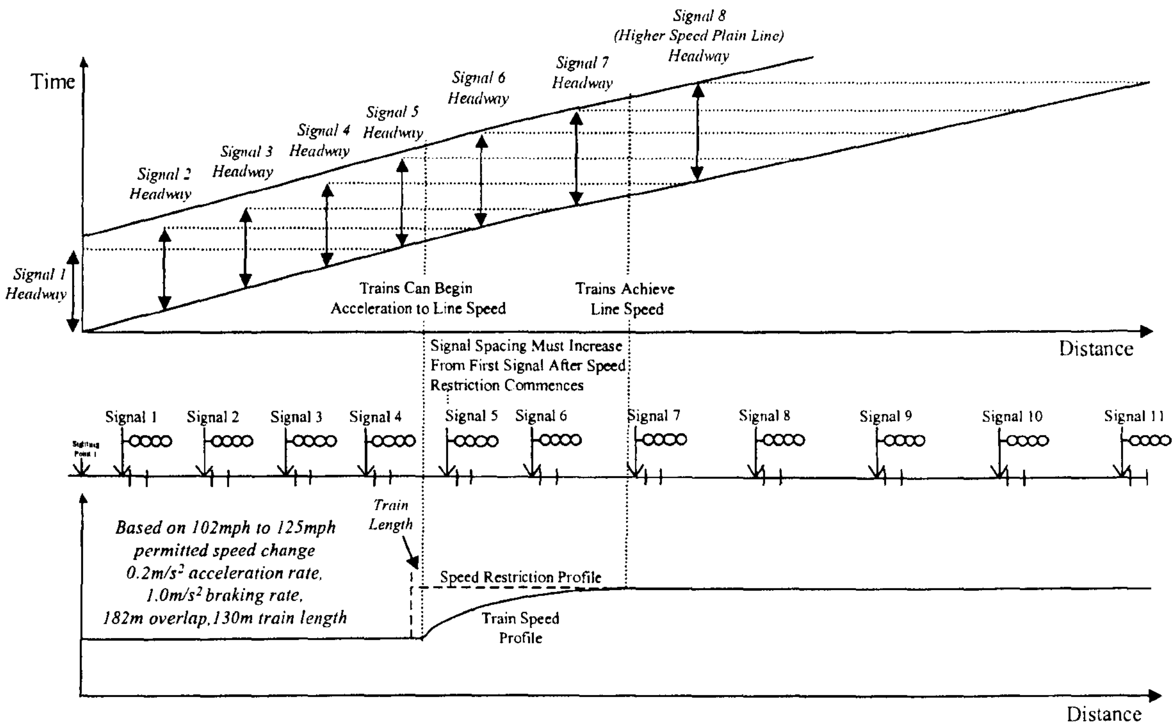


Figure 7-11: Headway for Increasing Speed Restriction with Signal Spacing Determined by Achievable Speeds

When the signal spacing and location of the speed restriction are known, the headway of a speed restriction can be treated in the same way as a station stop for consecutive stopping trains. For example, if the case of Figure 7-9 is taken, the critical headway has already been shown to be for the first signal within the area of the restriction and the headway distance will be given by:

$$H_{n(Higher\,Speed\,Restriction)} = \left[\frac{(n-1)V_R^2}{(n-2)2b}\right] + S + O + L \qquad \text{Equation 7-31}$$

Since the restriction commences before the signal, the exact location of the restriction will determine whether the sighting point is within the restriction or before it. If it is after the sighting point, the headway time will be given by:

Cruising time at lower speed + acceleration time + Cruising time at higher speed Equation 7-32

If the restriction commences at or before the sighting point, there will be no cruising time at the lower speed within the headway distance (and the acceleration time may commence from a higher speed than the lower restriction speed).

Where the area of line that could be safely traversed at a higher PSR is too short to make the adoption of increased signal separations worthwhile, the slower speeds of travel of the surrounding PSR sections would be maintained throughout the area.

7.3.5.2 TIGHTENING SPEED RESTRICTIONS- FIXED BLOCK SIGNALLING

If trains approach a PSR area at a higher speed, safety constraints require the braking distance provided from the signal before the restriction to be sufficient for that higher speed. This means that the separation of the first few signals in the restricted area must remain at that required by the higher permitted speed, even though trains must travel between them at the new, lower, PSR speed. This has a significant impact on achievable headway, as represented in Figure 7-12.

In the case represented in Figure 7-12, the longest headway time is that of the signal before the speed restriction commences.

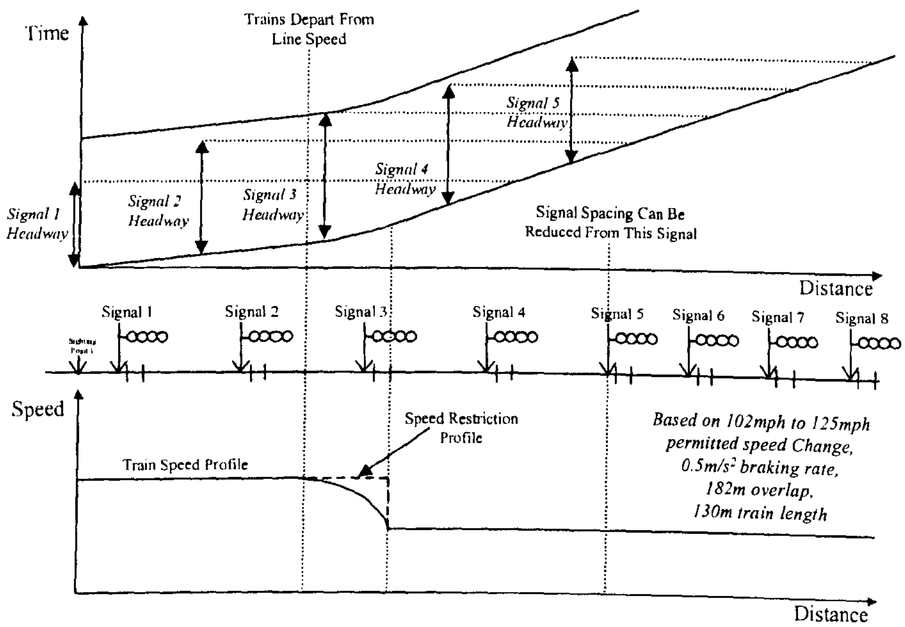


Figure 7-12: Headway for Decreasing Speed Restriction

As with a higher speed restriction, when the signal spacing and location of the speed restriction are known, the headway of a lower speed restriction can be treated in the same way as a station stop for consecutive stopping trains. The critical headway has been shown already to be for the signal before the restriction commences. Therefore, the headway distance will be given by:

$$H_{n(Lower\,Speed\,Restriction)} = \left[\frac{V^2}{2b} + \frac{V_R^2}{(n-2)2b}\right] + S + O + L \qquad \text{Equation 7-33}$$

Since the restriction commences after the signal, the exact location of the restriction will determine whether the sighting point is before or after the point at which braking to the restriction speed must commence. If braking must commence after the sighting point, the headway time will be given by:

Cruising time at higher speed + deceleration time + Cruising time at lower speed Equation 7-34

If the braking must commence at or before the sighting point, there will be no cruising time at the higher speed within the headway distance (and the deceleration time may commence from a lower speed than the higher restriction speed).

Where the area of reduced PSR is too short to enable adoption of reduced signal separations, the slower speeds of travel on approach to, through and immediately after the PSR extend the headway time, as represented in Figure 7-13.

The longest headway time in this example can again be seen to be that of the signal before the speed restriction commences.

If braking to the restriction commences after the sighting point of signal 3, the headway time will be given by:

Cruising time at Initial PSR + braking time to lower PSR + Cruising time at lower PSR + acceleration time back to higher PSR + Cruising time at higher PSR Equation 7-35

Once again, if braking to the restriction commences at or before the sighting point, there will be no cruising time at the higher speed within the headway distance (and the deceleration time may commence from a lower speed than the higher restriction speed).

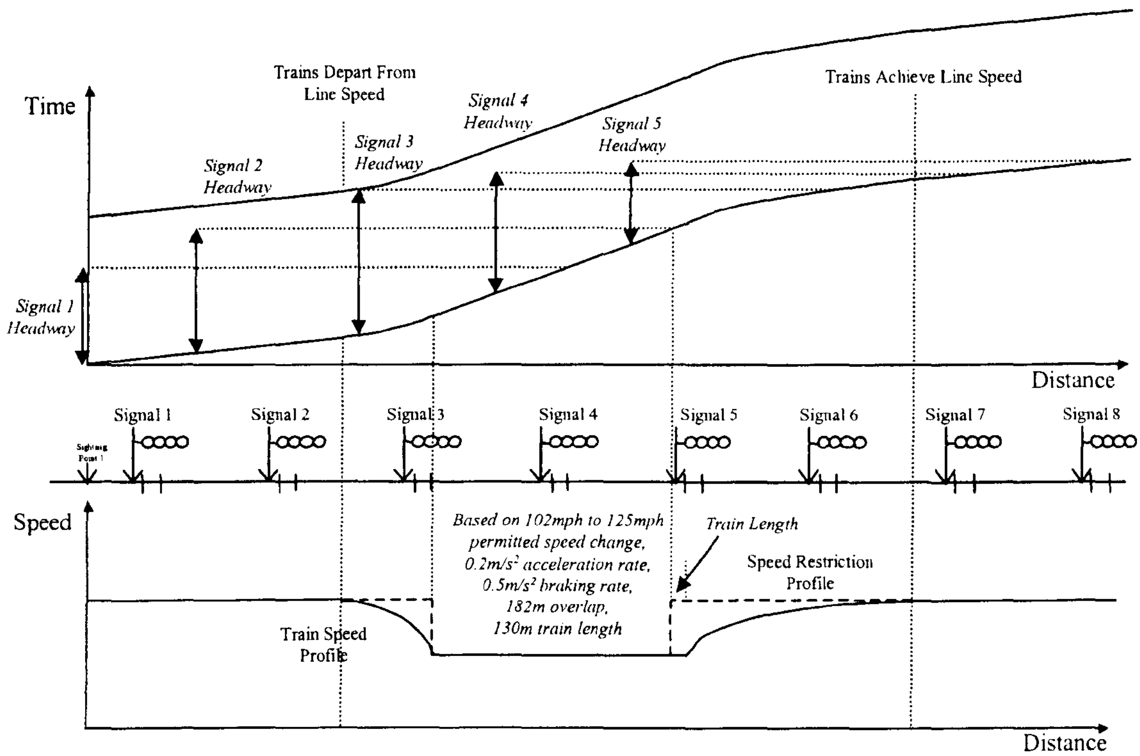


Figure 7-13: Headway for Short Speed Restriction

One final scenario remains to be considered in the topic of lower speed restrictions. This is the case of the restriction being longer than that considered in Figure 7-13, but for which no change in signal spacing is executed. This scenario is illustrated in Figure 7-14. Since the braking distance from signal 13 must be adequate to ensure that a train can stop at signal 5 from an approach at the higher speed restriction applicable at signal 3, it is only the block section between signals 5 and 6 that could be reduced in this scenario. If, however, it is not reduced, the most critical signal for headway purposes becomes the first signal within the restriction (signal 4) rather than the signal before the restriction. This is because the worst-case signal headway time in any scenario is determined by a combination of the headway distance and train's run-speed profile through the area. As trains' braking rates are generally higher than their acceleration rates, if the signal spacings are not reduced, any period of re-acceleration following a speed restriction will be more significant than the period of braking to it. The time spent at the restricted speed will also be more significant than either the acceleration or braking times.

The permutations and combinations that could occur in arrangement of speed restrictions make it impractical to develop detailed analytical equations for each possible case within this thesis. However, it is hoped that the general equations outlined here would be sufficient to enable the analysis of any specific case required.

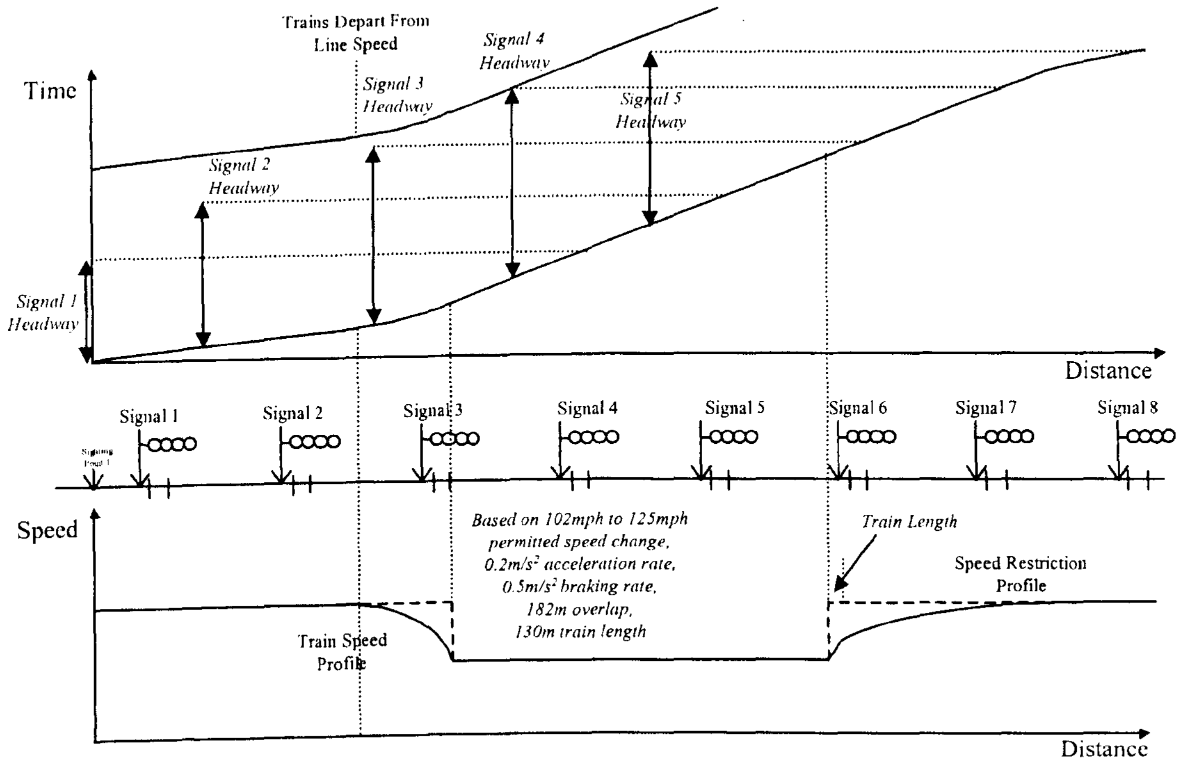


Figure 7-14: Headway for Long Speed Restriction (no change in signal spacing)

7.3.6 *JUNCTIONS*

The turnout speed at junctions is generally constrained due to the mechanical risk of damaging the point work or derailing the train. Trains passing through a junction on the straight through route are typically able to do so at the speed of any prevailing restrictions on approach to and exit from the junction area. Trains travelling along a diverging or converging route are usually further constrained by the junction turnout speed limit.

Where consecutive trains pass through a junction from and to the same line, the effect of the junction on headway is the same as that of a speed restriction (see section 7.3.4). Therefore, this situation will not be considered further here.

Where trains pass through the junction from or to different lines, the headway time between those trains will be affected by resetting of the rout between the trains.

7.3.6.1 *DIVERGING JUNCTIONS*

There are two possible train pattern scenarios for diverging junctions following different routes:

1. The first train following the straight through (un-restricted) route whilst the second train follows the diverging (speed restricted) route. The first train only needs to clear the junction (with some margin for delay in re-setting the route) before the following train can receive a proceed aspect at the first junction protection signal. The headway distance is, therefore, shorter than the train following headway distance, yet the train proceeds through the whole distance at the maximum permitted speed. This means that for all realistic conditions:
   - The headway time for this scenario will be less than the train following headway time for two trains travelling along either the straight through or diverging routes;

- The junction does not constrain the arrival or departing interval between trains.

  Since trains can only arrive at a junction with at least the train following headway time separation, the reduced headway cannot be utilised.

2. The first train follow the diverging (speed restricted) route whilst the second train follows the straight through (un restricted) route. As soon as the first train has cleared the junction, the route behind it can be released and reset for the following train. The following train can proceed unrestricted only if it arrives at the sighting point of the first junction protection signal after the route has been reset (see Figure 7-15). The headway time for this scenario will be less than that of two consecutive diverging trains, but longer than that of two consecutive trains on the straight through line. Therefore, it will affect the achievable capacity of the lines that the junction connects.

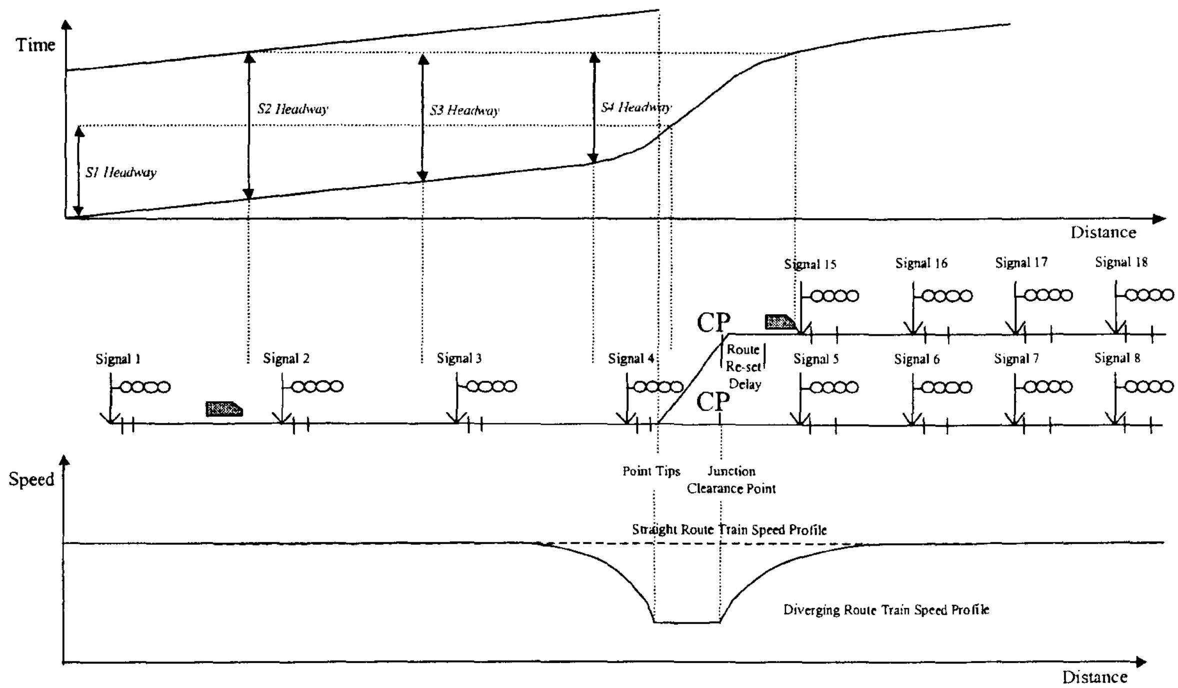


Figure 7-15: Headway for Diverging Followed by Straight Through Trains

7.3.6.2 CONVERGING JUNCTIONS

There are also two possible train pattern scenarios for converging junctions following different routes:

1. The first train following the straight through (un-restricted) route whilst the second train follows the converging (speed restricted) route. If unrestricted by signal aspects, the following train will gain on the lead train up until the point at which the lead train reaches the maximum permitted speed after the junction. In order to maintain headway separation at all times, this means that the following train cannot simply arrive at the sighting point of the junction protection signal as it turns green. Its arrival must be delayed until the first train is far enough ahead to ensure that it will have achieved the maximum permitted speed before the following train reaches a location the train following headway distance behind it. This means that the junction constrains the arrival interval between trains, but will not constrain their departing interval. This is represented in Figure 7-16;
2. The first train follows the converging (speed restricted) route whilst the second train follows the straight through (un-restricted) route. In this case, to maintain headway, the

following train must arrive at the sighting point of the junction protection signal as it turns green. It will then fall behind the first train until such time as it attains the maximum permitted speed after the junction. This means that the junction does not constrain the arrival interval between trains, but will constrain their departing interval. This is represented in Figure 7-17.

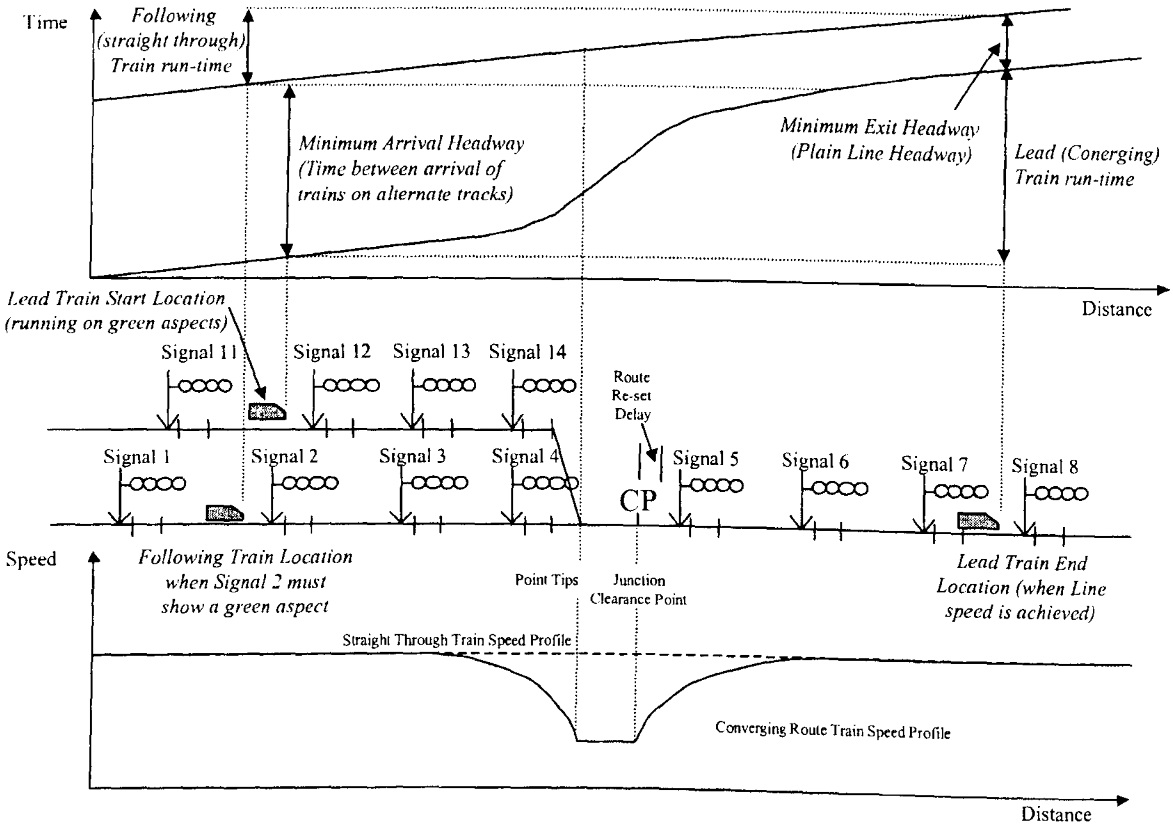


Figure 7-16: Headway for Converging followed by Straight Through Train

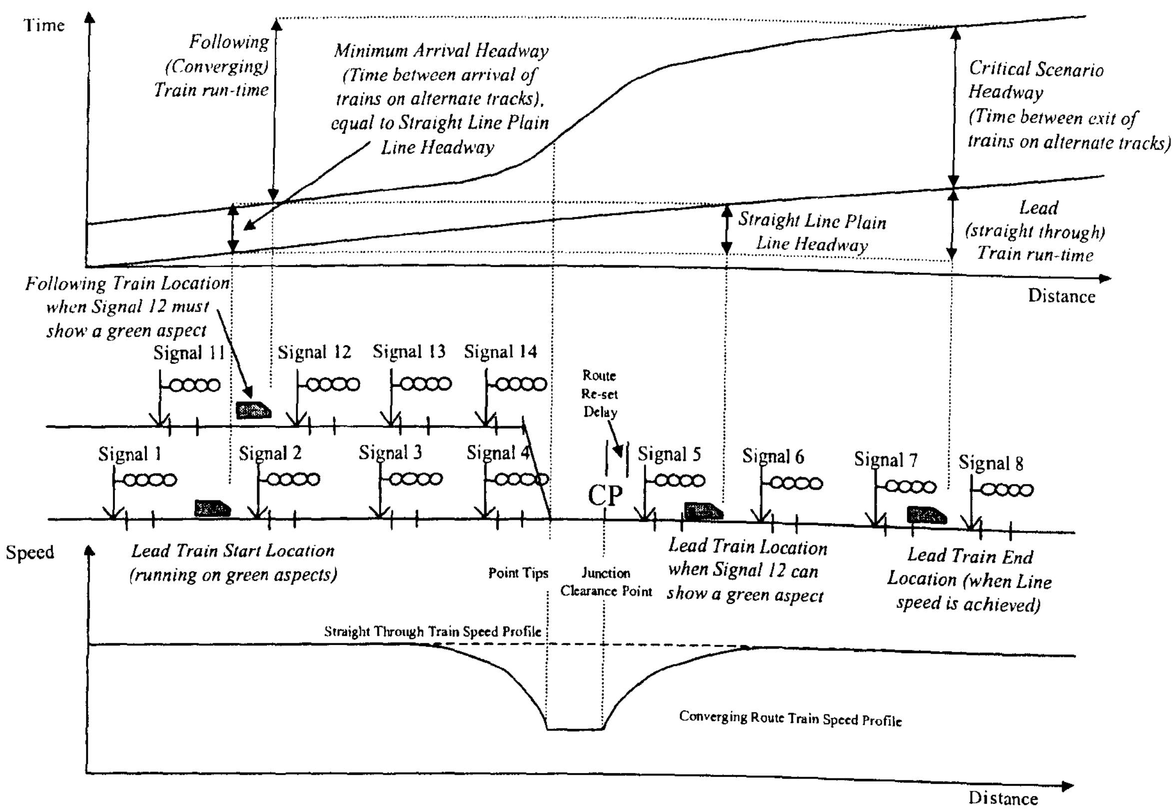


Figure 7-17: Headway for Straight Through followed by Converging Train

The headway of a converging junction is constrained as much by the junction speed and train performance as the location of signals. If the trains run-times are defined such that:

- The 'converging train run time' is given by: The time taken by a converging train to travel from the sighting point of the first junction protection signal on the converging line to the location at which it attains the maximum permitted speed after the junction;
- The 'straight through train' run time is given by: The time taken by the straight through train to travel from the sighting point of the first junction protection signal on the straight through line to the location at which a converging train attains the maximum permitted speed after the junction;

The critical headway time in both converging junction scenarios will be given by:

$$H_t = Train\ Following\ Headway + Converging\ Train\ Run\ Time - Straight\ Through\ Train\ Run\ Time \quad \text{Equation 7-36}$$

7.4 LUL 2 ASPECT POINT HEADWAYS

The demands for high train frequencies on LUL cannot be met by use of equi-spaced signals, particularly around station areas. LUL signalling is, therefore, designed with the signals located in optimal positions for minimising headway time. This can be achieved consistently for all trains, since LUL lines utilise homogeneous fleets of trains with similar stopping patterns.

The maximum line speeds used on LUL are relatively low, only reaching around 80km/h (50mph). This keeps sighting distances low and makes the use of 2 aspect signals appropriate.

It should be noted that, with the conventional trainstop based train protection used on most LUL lines, the overlap is required to be a full calculated braking distance from the maximum attainable speed at its start point. For the purpose of simplicity, this will be assumed to represent line speed. In practice, the short distances between many stations in the central London area mean that the maximum achievable train speed on approach to a station may be far lower than this. However, in that case the maximum achievable speed can be considered to be the effective line speed for that station area, making this a reasonable simplification.

7.4.1 ONE HOME (CONVENTIONAL)

For low (infrequent) headways, this typically requires a station starter located at the exit end of the station platform and a single 'Home' signal located an overlap distance before the entry end (LUL 1994, p30). This arrangement, represented in Figure 7-18, protects a train in the station from a following train.

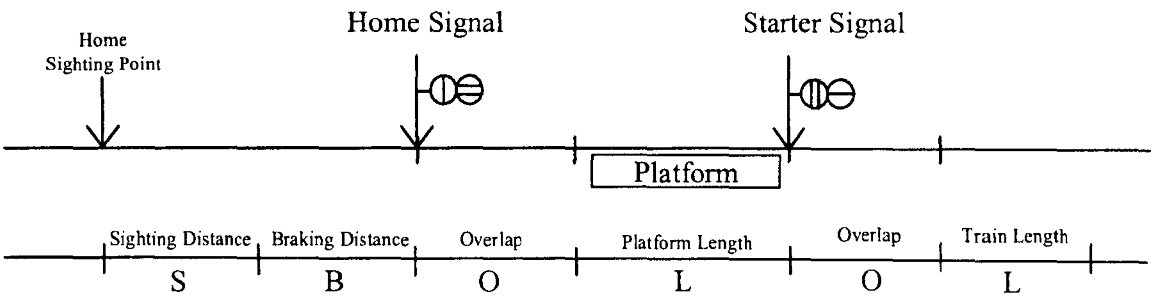


Figure 7-18: LUL 2 Aspect Signal Headways – One Home (Conventional)

The headway distances for this arrangement will be different for the home and starter signals. However, the critical factors are clearly the station approach and dwell time – making the headway distance of the home signal the one of most interest:

$$H_{(1H)} = S + B + 2O + 2L \qquad \text{Equation 7-37}$$

$$= \frac{V^2}{2b} + 2O + S + 2L \qquad \text{(Where b = deceleration rate)} \qquad \text{Equation 7-38}$$

In order to convert this to a headway time, the speed of the train during its passage through the route must be considered. During the sighting distance, the equivalent of an overlap distance and a train's length, the train can travel at line speed. It must then travel a braking distance in the process of braking to stop at the starter signal. After a dwell time, the train must then accelerate out of the platform and travel far enough to clear the overlap before the next train arrives. In doing this, depending on the line speed and lengths of the overlap and train, the train may reach line speed within the overlap or after clearing it. Therefore, the headway time can be derived as:

$$H_{t(1H)} = \frac{S+O+L}{V} + Braking\ Time + Dwell + Time\ To\ Clear\ Overlap \qquad \text{Equation 7-39}$$

Braking from line speed to rest takes V/b s, $V^2/2b$ m. The time to clear the overlap is given by:

$$t = \sqrt{\frac{2S}{a}} = \sqrt{\frac{2(O+L)}{a}} \quad \text{if line speed is not reached, or} \qquad \text{Equation 7-40}$$

$$t = \frac{V}{a} + \frac{O+L-\frac{V^2}{2a}}{V} = \frac{V}{2a} + \frac{O+L}{V} \quad \text{if line speed is reached} \qquad \text{Equation 7-41}$$

Where line speed is reached if: $V \leq \sqrt{2a(O+L)}$ Equation 7-42

In practice, sighting is usually taken to be a fixed time (S_t) rather than a distance (S). The equation for headway time of a single home signal is, therefore, given by:

$$H_{t(1H)} = S_t + \frac{O+L}{V} + \frac{V}{b} + Dwell + \begin{cases} \left\{\frac{V}{2a} + \frac{O+L}{V}\right\} & \text{For } V \leq \sqrt{2a(O+L)} \\ \left\{\sqrt{\frac{2(O+L)}{a}}\right\} & \text{For } V \geq \sqrt{2a(O+L)} \end{cases}$$

Equation 7-43

7.4.2 TWO HOMES (CONVENTIONAL)

If the headway that can be achieved by use of one home signal is insufficient, a second home (known as an inner home) can be located an overlap distance before the exit end of the platform (LUL 1994, p31). This allows the outer home to clear to a proceed aspect as soon as the preceding train has cleared the platform, as shown in Figure 7-19.

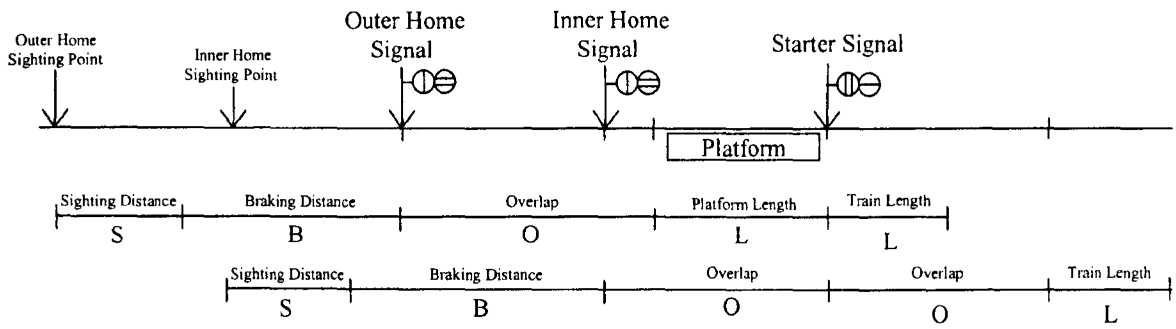


Figure 7-19: LUL 2 Aspect Signal Headways – Two Homes (Conventional)

Any train approaching the outer home can travel at line speed during the sighting distance, the equivalent of an overlap distance and a train's length. It must then travel a braking distance in the process of braking, in order to stop at the starter signal. After a dwell time, the train must then accelerate out of the platform before the next train arrives. Depending on the line speed, length and acceleration performance of the train, the train may reach line speed before leaving the platform, or after clearing it (although in practical scenarios it is likely to be after clearing it). Considering this description, the headway time can be derived as:

$$H_{t(2H,OuterH)} = \frac{S+O+L}{V} + Braking\ Time + Dwell + Time\ To\ Clear\ Platform \qquad \text{Equation 7-44}$$

Any train approaching the inner home can travel at line speed during the sighting distance and an overlap length. It must then travel a braking distance in the process of braking, in order to stop at the starter signal. After a dwell time, the train must then accelerate out of the platform and travel far enough to clear the overlap of the inner home before the next train arrives. The headway time can be derived as:

$$H_{t(2H,InnerH)} = \frac{S+O}{V} + Braking\ Time + Dwell + Time\ To\ Clear\ Overlap \qquad \text{Equation 7-45}$$

In this case, the time to clear the platform will be:

$$t = \sqrt{\frac{2L}{a}} \quad \text{for } V \geq \sqrt{2aL} \text{ or} \qquad \text{Equation 7-46}$$

$$t = \frac{V}{a} + \frac{L - \frac{V^2}{2a}}{V} = \frac{V}{2a} + \frac{L}{V} \quad \text{for } V \leq \sqrt{2aL} \qquad \text{Equation 7-47}$$

Taking sighting to be a fixed time (S_t), the equations for headway time in a two home arrangement are, therefore, given by:

$$H_{t(2H,OuterH)} = S_t + \frac{O+L}{V} + \frac{V}{b} + Dwell + \begin{cases} \left\{\frac{V}{2a} + \frac{L}{V}\right\} & \text{For } V \leq \sqrt{2aL} \\ \left\{\sqrt{\frac{2L}{a}}\right\} & \text{For } V \geq \sqrt{2aL} \end{cases}$$

Equation 7-48

$$H_{t(2H,InnerH)} = S_t + \frac{O}{V} + \frac{V}{b} + Dwell + \begin{cases} \left\{\frac{V}{2a} + \frac{O+L}{V}\right\} & \text{For } V \leq \sqrt{2a(O+L)} \\ \left\{\sqrt{\frac{2(O+L)}{a}}\right\} & \text{For } V \geq \sqrt{2a(O+L)} \end{cases}$$

Equation 7-49

It is the practice on LUL to assign a sighting time of 6 seconds to outer home signals, but a reduced sighting time of 2.5 seconds to subsequent signals on the station approach where provision of the full 6 seconds would limit achievable headway (LUL 1993, p3). This is considered acceptable on the grounds that sighting the outer home alerts the driver to the station approach, reducing the sighting time required to react to subsequent signals.

The limiting headway for the station stop will be the worst of the two headways. As both will always be an improvement on the one home arrangement, two home signals offer headway improvement at all speeds. This can be seen in Figure 7-21 in section 7.4.3.

7.4.3 THREE HOMES (CONVENTIONAL)

If the headway that can be achieved by use of two home signals is still insufficient, a third (intermediate) home can be added. This would usually be located an overlap distance before the middle of the platform (the exact location being determined by calculation to minimise headway at the designed operating speed)(LUL 1994, p32). This allows the outer home to clear to a proceed aspect whilst a train is still departing from the platform, as shown in Figure 7-20.

Following the same approach as before, the headway times for three homes can be derived as:

$$H_{t(3H,OuterH)} = \frac{S+O+L}{V} + Braking\ Time + Dwell + Time\ To\ Clear\ Middle\ Of\ Platform \quad \text{Equation 7-50}$$

$$H_{t(3H,IntermediaeH)} = \frac{S+O+0.5L}{V} + BrakingTime + Dwell + TimeToClearPlatform \quad \text{Equation 7-51}$$

$$H_{t(3H,InnerH)} = \frac{S+O}{V} + Braking\,Time + Dwell + Time\,To\,Clear\,Overlap \quad \text{Equation 7-52}$$

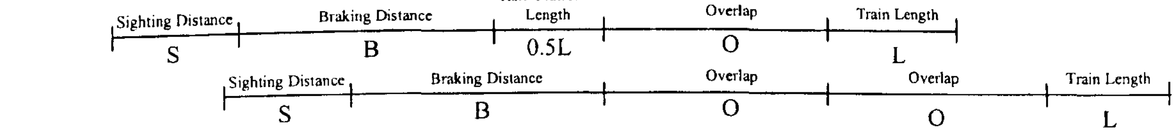


Figure 7-20: LUL 2 Aspect Signal Headways – Three Homes (Conventional)

Once again taking sighting to be a fixed time (S_t), the equations for headway time in a three home arrangement are, therefore, given by:

$$H_{t(3H,OuterH)} = S_t + \frac{O+L}{V} + \frac{V}{b} + Dwell + \begin{cases} \left\{ \dfrac{V}{2a} + \dfrac{0.5L}{V} \right\} & \text{For } V \le \sqrt{aL} \\ \left\{ \sqrt{\dfrac{L}{a}} \right\} & \text{For } V \ge \sqrt{aL} \end{cases}$$

Equation 7-53

$$H_{t(3H,IntermediateH)} = S_t + \frac{O+0.5L}{V} + \frac{V}{b} + Dwell + \begin{cases} \left\{ \frac{V}{2a} + \frac{L}{V} \right\} & \text{For } V \le \sqrt{2aL} \\ \left\{ \sqrt{\frac{2L}{a}} \right\} & \text{For } V \ge \sqrt{2aL} \end{cases}$$

Equation 7-54

$$H_{t(3H,Inner\,H)} = S_t + \frac{O}{V} + \frac{V}{b} + Dwell + \begin{cases} \left\{ \frac{V}{2a} + \frac{O+L}{V} \right\} & \text{For } V \le \sqrt{2a(O+L)} \\ \left\{ \sqrt{\frac{2(O+L)}{a}} \right\} & \text{For } V \ge \sqrt{2a(O+L)} \end{cases}$$

Equation 7-55

The limiting headway for the station stop will be the worst of these three headways. Whilst all three will always be an improvement on the one home arrangement, the inner home headway is the same for both two and three home signal arrangements. Providing three home signals therefore offers no headway improvement over two homes at speeds for which the inner home signal is the most restrictive. This can be seen in Figure 7-21.

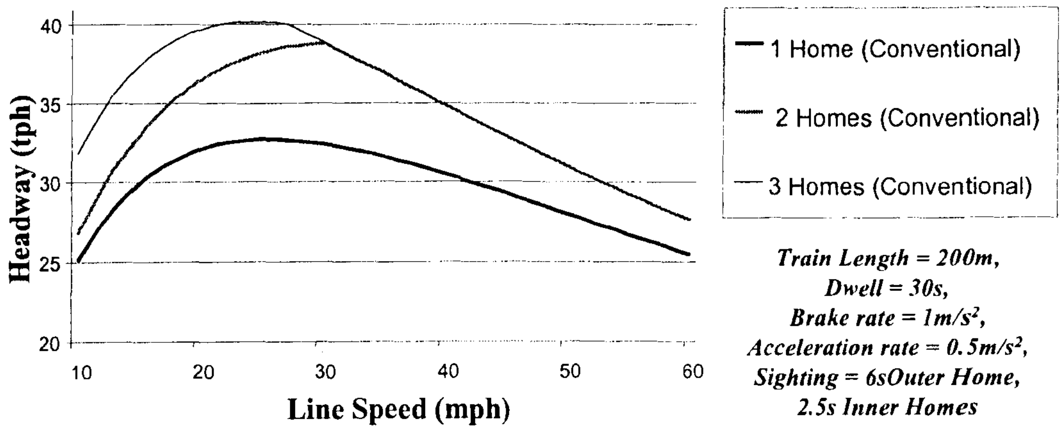


Figure 7-21: Conventional LUL Headways for 1, 2 and 3 Home Signals

Further reductions in limiting headways can be gained by the addition of further intermediate home signals. However, each addition offers diminishing returns and comes at a cost.

7.5 USING A BLOCK SECTION FOR THE OVERLAP

If the signalling of a line is arranged on the basis of equi-spaced fixed block signalling, the overlap need not be either a calculated braking distance (as used conventionally for LUL trainstop based signalling) or a nominal 183m (as used on most UK main line railways). It can, instead, be nominally assigned as the block section following the red aspect. Examples of this approach being adopted in the UK are the mass transit Central Line signalling system on LUL and the main line Channel Tunnel Rail Link (which utilises TVM430 signalling). In theory, a system can be developed on this basis to utilise any number of 'aspects' greater than 2.

7.5.1 TRAIN FOLLOWING HEADWAYS

As with conventional UK fixed block signalling, the minimum train following headway time will be achieved when the average signal spacing is given by:

$$d = \frac{B}{n-2}$$ Equation 3-9

and B is given by:

$$B = \frac{V^2}{2b}$$ (where b = braking deceleration rate; V = line speed) Equation 7-1

Since one aspect must be reserved for indicating all clear (equivalent to a Green aspect in conventional 2, 3 or 4 aspect signalling), two trains operating at the theoretical minimum train following headway with a block section for the overlap will always be separated by at least two block sections more than the braking distance, the sighting distance and the lead train's length. Their train following headway distance will be given by:

$$H_n = \frac{V^2 n}{2b(n-2)} + S + L$$ (where n>2) Equation 7-56

Or the headway time by:

$$Ht_n = S_t + \frac{Vn}{2b(n-2)} + \frac{L}{V}$$ (where n>2) Equation 7-57

This is represented (together with comparable results for a nominal 182m overlap system) in the left hand graph of Figure 7-22. The right hand graph in the same figure represents the difference in train following headway with block section overlaps and nominal 182m overlaps when they are implemented for signalling systems with the same number of aspects (the results for 3 to 6 aspects being shown). The difference between train following headways with block section overlaps and nominal 182m overlaps will be discussed in more detail in section 7.5.2.

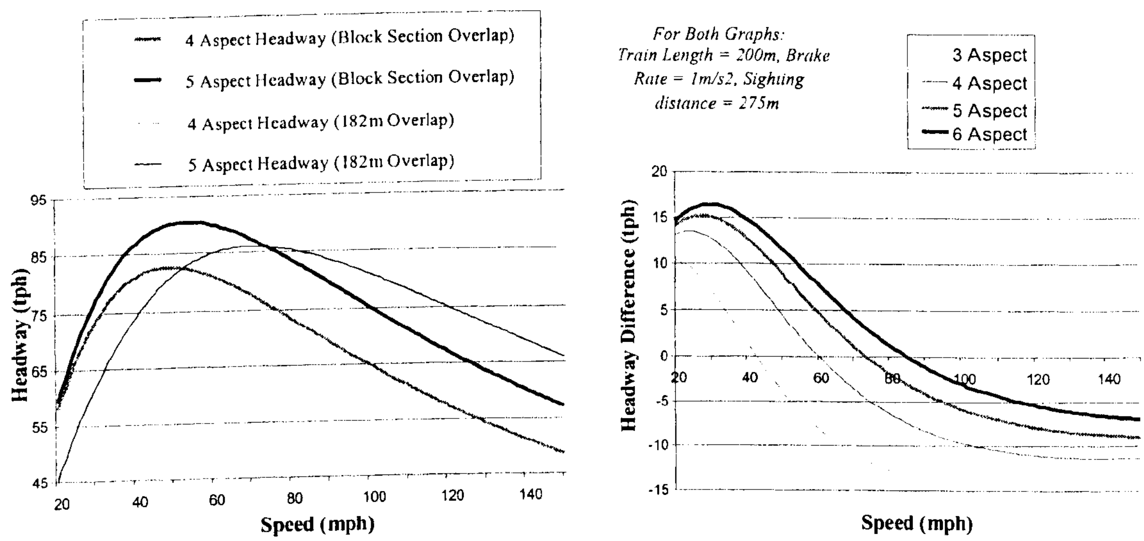


Figure 7-22: Comparison of Train Following Headway for use of a 182m Overlap and a Block Section Overlap

7.5.2 POINT HEADWAYS

The point headway time for consecutive stopping trains was derived in section 7.3.1 as:

$$H_{m(Stopping)} = \frac{S}{V} + \frac{V(2n-3)}{b(2n-4)} + Dwell + \begin{cases} \sqrt{\dfrac{2(O+L)}{a}} & \text{For } V \geq \sqrt{2a(O+L)} \\ \dfrac{V}{2a} + \dfrac{O+L}{V} & \text{For } V \leq \sqrt{2a(O+L)} \end{cases}$$ (for n > 2)

Equation 7-19

If the overlap is a block section rather than a nominal distance this equation still stands, with the value of the overlap distance given by:

$$O = \frac{V^2}{2b(n-2)}$$ Equation 7-58

The headways obtained by use of block section and nominal 182m overlap systems can be seen in Figure 7-23. Both 4 and 6 aspect signalling are shown, representing the typical number of aspects in a conventionally UK main line signalling system (using nominal overlaps) and TVM-430 (using block section overlaps).

It can be seen that the most significant factor in determining the difference in theoretical headway between signalling systems that utilise both block section overlaps and nominal 182m overlaps is the line speed. The number of aspects has a relatively small effect.

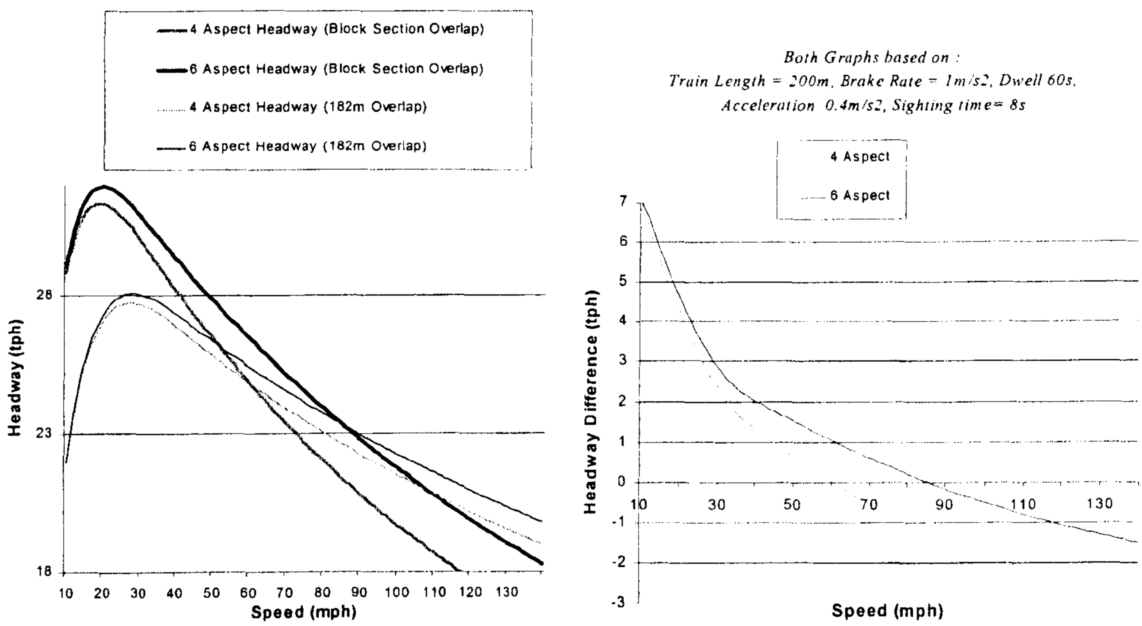


Figure 7-23: Comparison of Point Headway for use of a Nominal 182m Overlap and a Block Section Overlap

At low speeds (when the signal separation is required to be less than 182m) Figure 7-23 shows that the use of overlaps equal to the signal separation (block section) offers significantly improved headway over that of nominal 182m overlaps. However, at low speeds a more traditional signalling design would use a nominal overlap of less than 182m. Therefore, this apparent improvement in capacity would not be fully achieved in practice. An example of this can be seen in the Network Rail standards for provision of overlaps, which permit reduced overlap lengths at speeds below 60mph, with as little as 45m being required for 15mph operation (Marks 2000, p5).

A comparison of the theoretical headway for block section and Network Rail standard overlaps is shown in Figure 7-24. With line speeds below about 70mph, the scale of capacity improvements offered by block section overlaps is significantly reduced, but for a 6-aspect system still appreciable (in the order of 0.75tph). For speeds between 70 and 100mph the difference is marginal for a 6-aspect system, but shows a significant benefit from the use of nominal overlaps (in the region of 1 to 2 tph). This effect is even more pronounced at higher speeds. The use of block section overlaps in a 6-aspect arrangement imposing a loss of around

1.5tph at 140mph, whilst in a 3-aspect arrangement it would impose a loss of around 3tph at the same speed.

At the low line speeds typical on metro railways, the use of a block section overlap appears to offer the potential for some capacity benefit whilst also reducing the required complexity of track circuit arrangements. This approach, therefore, offers advantages to such operation. However, the use of block section overlaps would appear to be unsuited to high-speed operation.

The use of a block section as an overlap is very similar in principle to the more usual UK practice of a nominal overlap. It is not, therefore, proposed to pursue analysis of it any further.

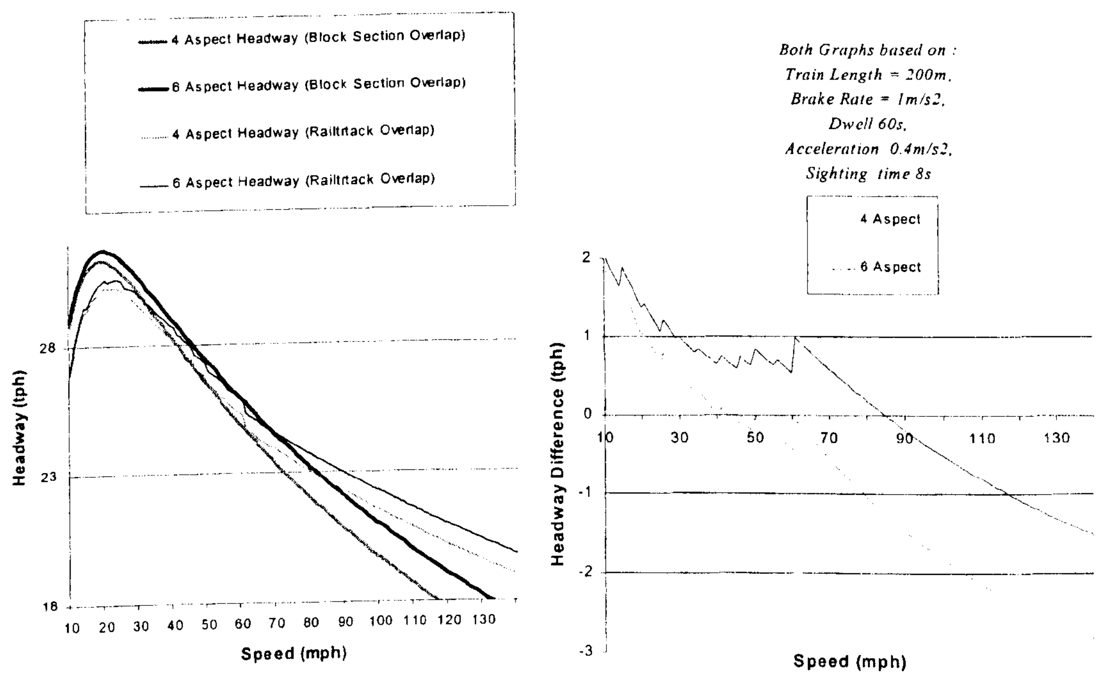


Figure 7-24: Comparison of Point Headway for use of a Network Rail Standard Overlap Lengths and a Block Section Overlap

7.6 SUMMARY

On first consideration, the operation of trains at speeds lower that implied by the signalling design appears to reduce the capacity of the line significantly. In practice, however, it was shown in section 7.2.1 that the reduction in braking distance that accompanies lower speed operation could permit a smaller number of block sections to be maintained between trains without impacting safety. This in turn means that the designed capacity could be maintained (or at least reduced by a lesser amount) at lower speeds of operation. In some cases, a higher capacity than the designed capacity could even be achieved – particularly with moving block operation. In light of this, both actual speeds of operation and the implications that they place on safe operating practice must be considered when determining the capacity of a given line and considering the impact of operation at reduced speeds.

A further significant point noted in section 7.2.1 is that the moving block signalling not only offers improved train following headway performance at reduced speeds when compared to fixed block multiple aspect systems, but that the improvement is more noticeable at higher designed line speeds. This suggests that moving block operation offers greater benefits to high-speed main lines with intermixed traffic types, or during perturbations that cause slower speed operation, than to lower speed metro type services.

In section 7.3 it was found that the slower speeds of operation required around stations, junctions and speed restrictions can significantly impact the theoretical headway. In station and junction areas, the increase in headway time acts as a major constraint on theoretical headway through the line. Merely increasing the train following headway may not, therefore, provide useable capacity. It is the pinch points of the stations and junctions that require the most attention if overall capacity is to be increased.

The equations derived within this chapter enable a comparison of the capacity impact arising from selection of different train separation strategies. This is represented in Figure 7-25 for 4 aspect signalling (both with nominal 182m and block section overlaps), moving block and conventional LUL 2 aspect signalling with 3 home signals.

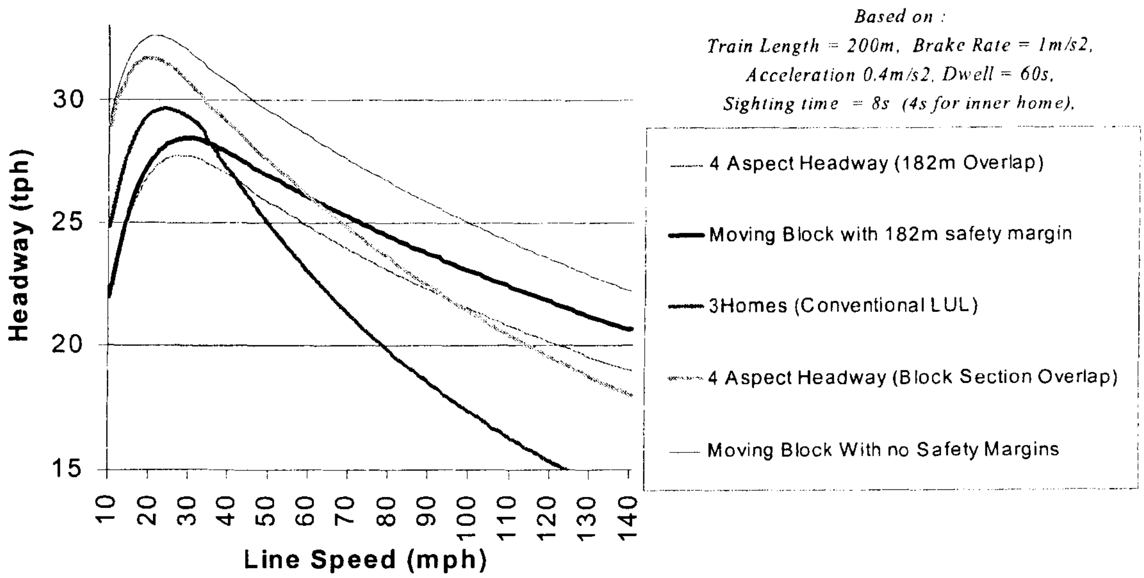


Figure 7-25: Comparison of Point Headway for 4 Aspect and LUL 3 Home Signalling

It should be noted that different train separation strategies result in significantly different theoretical headways, with the optimal system being determined by the maximum speed of operation.

8 PRACTICAL CONTROL SYSTEM CAPACITY

8.1 INTRODUCTION

In the analysis of control system capacity contained within chapters 3, 4 and 7, the author developed equations for theoretical headway evaluation. That is, the headway that could be achieved by operation of trains over a single uni-directionally operated line equipped with in-line stations, ideally spaced signals, a signalling system that imposed no delays and using the best possible train speed profile at all times.

In any practical railway, the actual achieved headway will be influenced by additional factors, including:

- Variations in signal spacing;
- Delays imposed by the signalling system;
- Train performance (acceleration and braking rates which may be variable);
- Operator behaviour;
- The provision of additional off-line platforms
- The need for termini at the ends of the line.

In addition to these factors, whether on a single or multiple line railway, interactions can be expected between movements planned for services heading in different directions. Such conflicts will also impact the capacity of the line.

In this chapter, the author will further develop the headway equations derived in chapter 7 in order to account for the practicalities of operating a real railway. The focus of the chapter will be on fixed block multiple aspect signalling with a nominal (standard) overlap.

8.2 PRACTICAL SIGNALLING APPLICATION

Whenever headway has been discussed in previous sections, it has been assumed that signals will be placed at braking distance separation for sections of 3-aspect signalling, at half braking distance separation for 4-aspect signalling, etc. However, for optimum stopping headway time through a station with 4-aspect signalling, the innermost home signal must be located a distance equal to the train length and overlap before the station starter. The intermediate home must be located braking distance from the starter signal and the outer home at braking distance from the inner most home signal (Anon. 1967, p607). This arrangement is represented in Figure 8-1.

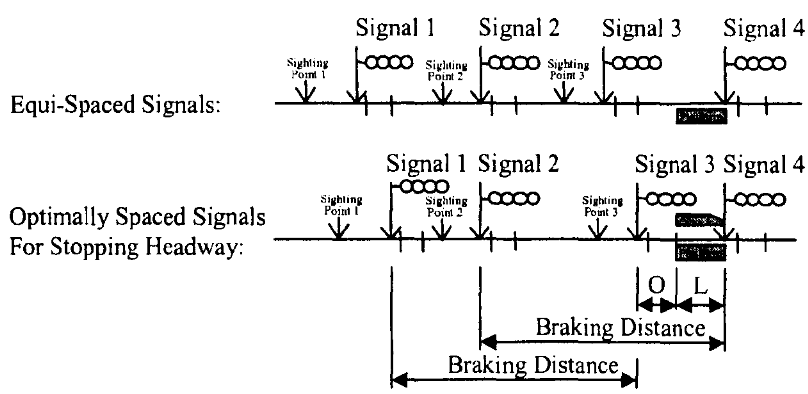


Figure 8-1: Optimal 4-Aspect Signalling Layout for Stopping Headway

As with the LUL approach to signalling layout design (discussed in section 7.4), this arrangement protects a train in the station from a following train, whilst allowing the following train to approach as closely as possible before encountering the outer home signal. In order to locate signals at these optimum separations, they would clearly not be equi-spaced. The starter and inner home signal would be separated by less than half braking distance, as would the outer and intermediate home signals. However, the intermediate and inner home signals would be separated by more than half braking distance. The safety of train movements can still be assured by this arrangement, since full braking distance is maintained between the outer and inner homes and between the intermediate home and station starter.

By optimal spacing of signals, within the constraint of alternate signals remaining at full braking distance separation, the headway of a 4-aspect signalling system could theoretically be improved over the case of assumed equal spacing between signals. Similar optimisation could also be applied to systems with more than 4 aspects.

Unfortunately, on a real railway, optimal location of signals is rarely possible. The location of signals may be constrained by factors such as the location of stations (including the separation between them), point work, tunnels and viaducts or the need for clear sighting of the signals' aspects for a sufficiently long time on approach to them. As a result, signals are mostly over-braked (located with a separation that exceeds the requirements for braking).

It is generally considered that signal separations should not be more than 50% above the braking distance, due to the impact that longer distances would have on headway and driver concentration (Nock 1980, p6). Significant variation in signal separation also forces drivers to rely extensively on their route knowledge and can make it difficult for them to judge braking distances. This makes the risk of signal overruns higher (Fleming 2000, p5). In order to control both the capacity and risk implications of extended signal separations, Network Rail require the differences in excess spacing between consecutive signals to be kept to a minimum and also apply constraints on the maximum permitted signal spacing:

- Up to 33% excess without risk assessment
- 34 to 100% excess for lines with signal spacing in excess of 500m, subject to satisfactory overrun risk assessment
- Up to 1000m for lines with signal spacing of less than 500m, subject to satisfactory overrun risk assessment

  (Flemming 2000, p5).

Over-braked signal separations increase the headway distance, and thus the headway time, reducing the potential throughput of trains.

The practical constraints imposed on signal location mean that there can be no guarantee that separation distances will comply with some pre-defined formula. The actual distances that arise from the process of signalling layout design must thus be accounted for in determining theoretical headway.

As already discussed in section 7.5.2, the length of overlaps on Network Rail infrastructure are nominally 180m, but at low speeds the overlap length can be reduced to as little as 45m at 15mph. Reducing the length of the overlap offers improvements in headway at the expense of an increased probability of trains overrunning the provided overlap in the case of driver error, poor adhesion or failure of the train braking system. The allowed reductions in overlap length are, therefore, only applied in practice where the headway improvement offered is critical to achieving the desired capacity and a risk assessment shows that it is acceptable to do so (Marks 2000, pp5 & 8). This means that there can also be no guarantee that overlap lengths will comply with some generic rule.

In addition to this, whilst the author's deliberations have so far assumed braking distances to be given by $V^2/2b$, where the brake rate (b) is a constant value, they would be longer than this in practice. The equations used by Network Rail to determine braking distances include allowances for:

- The stopping distance from a given speed under good rail conditions (determined by practical tests);
- The time taken (and therefore distance travelled) between brake demand and full brake application. This includes an allowance for a delay to start of brake build up and for a gradual increase in brake force once the brakes do apply;
- Variation in the brake rate as a function of speed;
- The acceleration and deceleration effects of gravity where gradients are not level;
- Speedometer and other measurement errors (for which an allowance of 10% is added to the calculated stopping distance);
- Areas of poor adhesion, where the assumed braking rates will not be achieved (a further allowance of 10% is added to the calculated stopping distance, subject to a minimum amount of 10m, to allow for this).

(Fleming 2000, p25)

The equivalent calculation used by LUL in determining the sighting point of signals assumes that braking requires a distance given by $V^2/2b$. However, the allowance added to this for driver sighting means that accurate calculation of the braking distance is not critical. This is not the case during emergency braking after tripcock activation at a trainstop. In this circumstance LUL makes allowances for:

- The time taken to achieve full brake application;
- The assumed braking rate (with different rates assumed for tunnel and open sections, to reflect the different adhesion levels expected);
- The acceleration effects of gravity where gradients are not level;
- An allowance of un-known factors (such as wind effects);
- A 5% safety margin added to the total calculated distance.

(LUL 1993, pp5-6)

This results in an equation for the calculation of overlaps given by:

$$O = 3 + 1.05\left(1.5V + \frac{V^2}{2(b+0.1G)} + \frac{1.309V^4}{10^4}\right) \quad \text{(LUL 1993, p5)} \qquad \text{Equation 8-1}$$

Where the tripcock is 3m from the front of the train, G is the percentage gradient, V the maximum achievable speed at the start of the overlap (m/s) and b the braking rate (m/s^2)

As a third example, braking distances on New Jersey Transit are calculated allowing for 8s of free running and a 25% safety margin (Troup 2003, p4).

Accounting for factors such as these, realistic braking distances would be expected to be longer than $V^2/2b$. Signal separations would, therefore, also need to be longer than the theoretical minimum so far considered.

Practical headway calculation must account for these elements of signalling design. They must also account for the practicalities of operating signalling equipment, such as:

- Delays within the train detection equipment that result in occupation and clearance of track circuits not registering at the precise moment they occur (a particularly significant and variable factor in modern solid state interlockings);
- Delays within the signal aspect control circuitry that result in signals not clearing to less restrictive aspects at the precise moment such clearance becomes permissible in accordance with detected train positions;
- Delays in resetting the route between trains traversing junctions in different directions, including:
- A route release delay. Time for the train detection and interlocking to respond and release the route behind the first train, once the conditions required for its release have been met;
- A supervisory system delay. Time for the supervisory system (or signaller) to recognise that the route has been released and to output controls to the interlocking to set the route for the following train;
- A route set delay. Time for the interlocking to set the route for the following train;
- A point swing delay. Time for the points to move to their new position and for the completion of the movement to be detected by the interlocking;

8.2.1 FORWARD SPEED PROFILE

In the light of the factors discussed in section 8.2, analysis of practical headways becomes more complex than the theoretical scenarios discussed in chapter 8. This is represented for a 4-aspect section of plain line railway (that is, a section with no junctions) in Figure 8-2.

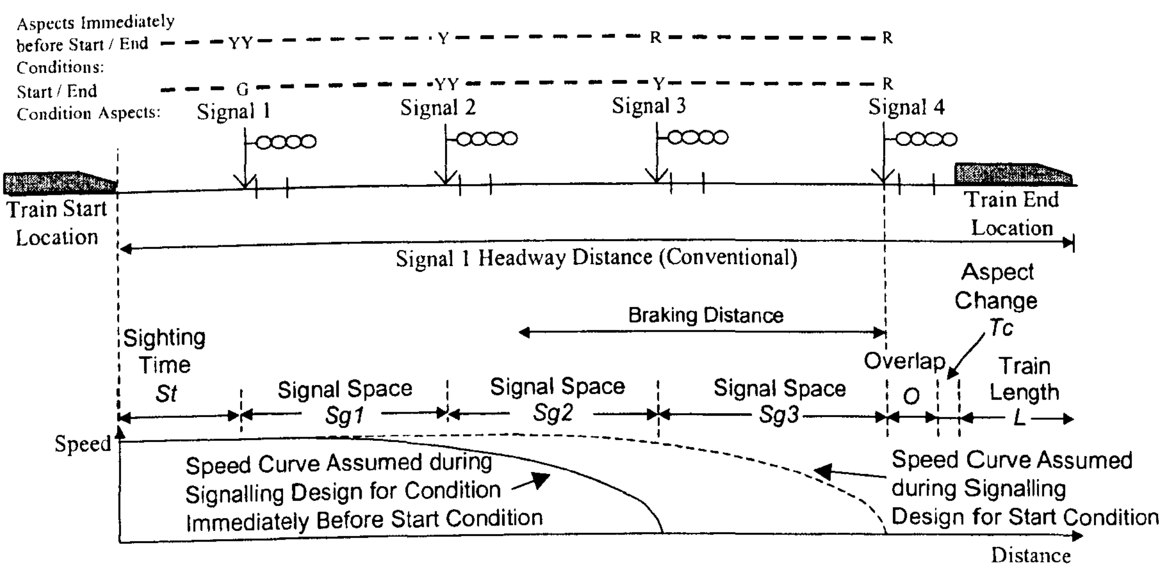


Figure 8-2: Conventional Plain Line Technical Headway

Therefore, assuming a constant value for the permitted speed, V, the theoretical headway time for trains operating over a section of 4 aspect signalling is given more accurately by:

$$Ht_4 = St + Tc + \frac{(Sg1 + Sg2 + Sg3 + O + L)}{V} \qquad \text{Equation 8-2}$$

Where: Tc is the time for signal 1 aspect to change once the conditions for it to do so have been physically satisfied. This includes any delay in detection of the conditions, time for the interlocking to react and for the aspect to subsequently change.

In practice, however, the permitted speed is unlikely to be constant and it is also unlikely that trains travelling through the section will always be able to do so at exactly the permitted speed. This is most obviously true where a station stop is required, but is also true where a speed restriction occurs. Whilst a change in permitted speed will occur as a step function, trains must reduce and/or increase speed gradually limited by the train's braking and traction performance and, thus, cannot follow the permitted speed profile exactly. Such deviations from the permitted speed must be included in any calculations to determine practical headway.

Accounting for this, the headway time could be represented as:

$$Ht_4 = St + Tc + \frac{(Sg1 + Sg2 + Sg3 + O + L)}{FSP} \qquad \text{Equation 8-3}$$

Where: FSP = Forward Speed Profile (the actual speed profile of the train travelling through the section of track being considered).

Unfortunately, as the FSP varies continuously, this equation is only a pseudo-quotient and the result cannot be calculated. However, another way of expressing Equation 8-3 would be:

$$Ht_4 = St + Tc + T_{FSP} \qquad \text{Equation 8-4}$$

Where: T_{FSP} is the time taken to travel the distance $Sg_1+Sg_2+Sg_3+O+L$ in accordance with the FSP.

This equation can be calculated by considering the relationship between the instantaneous speed (V_{act}) and location (x) within the FSP:

$$V_{act}(x) = \frac{dx}{dt} \quad or \quad 1 = \frac{1}{V_{act}(x)}\frac{dx}{dt} \qquad \text{Equation 8-5}$$

This function can be integrated with respect to time to give:

$$\int_0^t 1dt = t \quad = \int_0^t \frac{1}{V_{act}(x)}\frac{dx}{dt}dt = \int_0^t \frac{dx}{V_{act}(x)} \qquad \text{Equation 8-6}$$

If the train location when t equals 0 is defined to be location 'x_{start}' (the location of signal 1), and it takes a time T_{FSP} to travel the distance $Sg_1+Sg_2+Sg_3+O+L$ in accordance with the FSP, the integral becomes:

$$T_{FSP} = \int_{x_{start}}^{x_{start}+Sg1+Sg2+Sg3+O+L} \frac{dx}{V_{act}(x)} \qquad \text{Equation 8-7}$$

Applying this to Equation 8-4, the headway time is given by:

$$Ht_4 = St + Tc + \int_{x_{start}}^{x_{start}+Sg1+Sg2+Sg3+O+L} \frac{dx}{V_{act}(x)} \qquad \text{Equation 8-8}$$

This equation can be generalised to represent an 'n' aspect signalling system:

$$Ht_n = St + Tc + \int_{x_{start}}^{\left(x_{start}+O+L+\sum_{p=1}^{p=n-1} Sg_p\right)} \frac{dx}{V_{act}(x)} \qquad \text{Equation 8-9}$$

Whilst integrals such as this cannot be easily calculated, any integral can be approximated by a finite summation. In this case, the finite summation would involve adding together the times taken to make progressive steps of distance Δx, assuming the dynamic parameter $V_{act}(x)$ remains constant during each step. The result of the summation will represent a close approximation of the integral so long as the summation steps are kept small enough to make changes in $V_{act}(x)$ negligible over the duration of any individual step.

In this case, adopting a finite summation approximates the headway time to be:

$$Ht_n = St + Tc + \sum_{x_{start}}^{\left(x_{start}+O+L+\sum_{p=1}^{p=n-1} Sg_p\right)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation 8-10}$$

During any step within this finite summation, $V_{act}(x)$ can be determined by considering the acceleration applicable over the interval since the previous step.

This same approach can be adopted to analysing the time taken for a train to travel between two defined locations in accordance with the applicable forward speed profile, such that:

$$T_{FSP} = \sum_{x_{start}}^{(x_{end})} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation 8-11}$$

The use of this equation will become apparent through the remaining sections of this chapter.

8.3 DIVERGING JUNCTIONS

In practice, diverging junctions are rarely as straightforward as the description given in chapter 8. The need to ensure that drivers are aware of, and comply with, junction speed restrictions often leads to some form of approach control being applied to the junction signal. In such cases, the signal is held at a restrictive aspect, even if the conditions ahead permit a less restrictive one, until both the route indication and main signal are readable by the driver. The approach control is then released (referred to as approach release), so that the signal can display the aspect actually permitted by the conditions ahead. On Network Rail infrastructure, there are clearly defined criteria for the appropriate arrangement of junction signalling to be applied. These are:

- Where the difference in speed between the straight through and diverging route is less than 10mph no approach control is required (the driver is expected to recognise the need for speed reduction and implement speed reduction him/her self). Since speeds are set in 5mph increments, this in practice means a difference of not more than 5mph;
- Where a more significant speed differential exists, but all trains using the route can brake to the restriction in the distance between the signal and turnout, the junction protection signal may be approach released from yellow;
- Where approach control is desired, but release from yellow would be too restrictive, approach release from flashing yellows in rear may be provided;

- As an alternative to approach release from flashing yellow, a splitting distant signal may be used;
- Where the restriction for the diverging route is too great for approach release from yellow to be used, or approach release from flashing yellow / splitting distants have been used for another route from the signal, approach release from red may be used instead; (Woolford[1] 2002, pp30-1).

It was noted in section 7.3.6.1 that, if the first train travelling over a junction follows the diverging (speed restricted) route whilst the second train follows the straight through (un-restricted) route, the headway time might be longer than the plain line headway. With the introduction of approach control to a junction's signalling arrangements, a second scenario also becomes of interest. That is, the case of consecutive trains following the diverging (speed restricted) route. In such a case, the movement of both diverging trains will be restricted by the signalling arrangement, which may have the effect of reducing the headway impact imposed by the junction speed restriction.

8.3.1 DIVERGING FOLLOWED BY STRAIGHT THROUGH - NO APPROACH RELEASE

This situation, already described in section 7.3.6.1, is represented for 4-aspect signalling in Figure 8-3. As soon as the first (diverting) train has cleared the junction, the route behind it can be released and reset for the following train. With no approach release, headway will be maintained if the following train arrives at the sighting point of the first junction protection signal (signal 2 in Figure 8-3) as it clears to a green aspect in response to the completion of this reset. Since signals closer to the junction would also be released at the same point/time, they would have a smaller headway time.

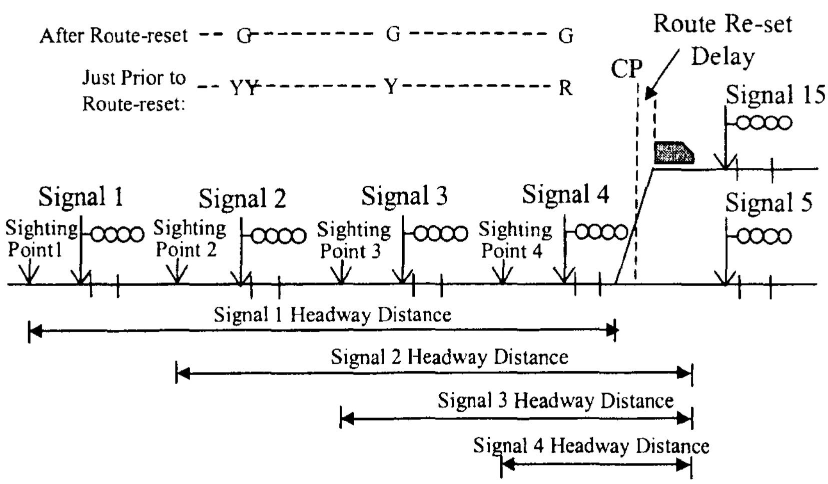


Figure 8-3: Headway for Diverging Train Followed by Straight Through Train, No Approach Release

The headway of the first junction protection signal would be given by:

$$Ht_{n(diverging\ junction-diverging\ /\ straight)} = St + Tc + Tr + \sum_{x_{start}}^{\left(x_{start}+Sj+L+\sum_{p=1}^{p=n-2} Sg_p\right)} \frac{\Delta x}{V_{act}(x)} \quad \text{Equation 8-12}$$

Where: Sj is the distance between the last junction protection signal and the junction clearance point;

Tr is the time taken for the route to reset behind the first train;

Tc is the time for the first junction protection signal's aspect to change once the conditions for it to do so have been physically satisfied.

Depending on the combination of forward speed profiles, junction location and route reset delays, the headway time imposed by the first junction protection signal may represent the critical headway for the junction area or may be shorter than that of the preceding signal (signal 1 in the example of Figure 8-2). The preceding signal headway can be calculated in accordance with equation 8-10 in section 8.2.1. The limiting headway of this scenario will then be the longer of the two calculated times.

8.3.2 DIVERGING FOLLOWED BY STRAIGHT THROUGH - APPROACH RELEASE FROM YELLOW

Where approach release from yellow is used, although the speed profile through the route does not require braking until the junction protection signal has been passed, the signal is held at yellow until the train reaches a point within which both the main aspect and the junction indication are readable. The approach release will then be initiated by either:

- Occupation of a track circuit commencing at the readable distance, or;
- Expiry of a timer started by occupation of a track circuit that commences before the readable distance. In this case the timer would be selected such that a train travelling over the track circuit at the maximum permitted speed would be at the readable distance when the timer expired.

For colour light signals and junction indicators, the readable distance is considered by Network Rail to be 800m. If an alphanumeric route is used, the readable distance is considered to be 250m (Woolford[(2)] 2002, pp30-31).

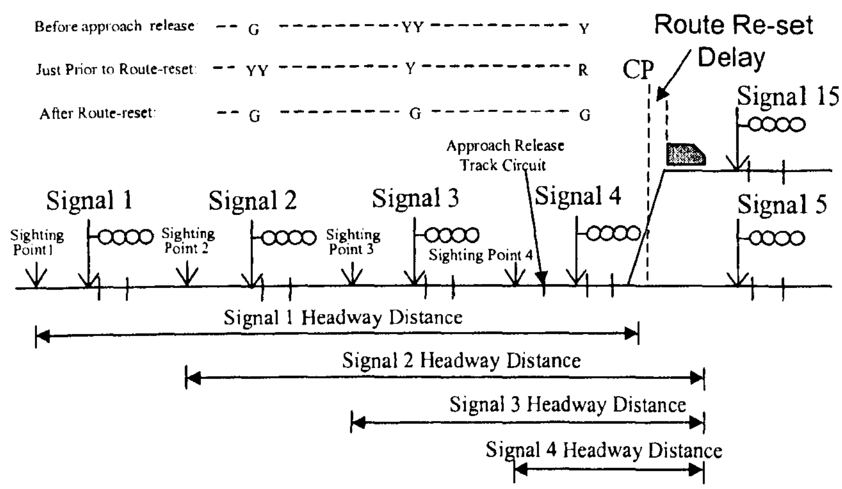


Figure 8-4: Conventional Headway for Diverging Train Followed by Straight Through Train, Approach Release from Yellow

As shown in Figure 8-4, the critical headway for this junction arrangement and service pattern will again be that for the first junction protection signal (unless the plain line headway on approach to the junction area is more restrictive – which becomes increasingly unlikely as the junction speed is reduced). It should be noted that, although the first (diverting train) will never approach signal 4 displaying an aspect less restrictive than yellow prior to the approach release conditions being fulfilled, this does not influence the headway distance. This is due to the need for the following train to receive green aspects if headway is to be maintained. As a consequence of this, the headway time will still be given by equation 8-12. However, it will be influenced by the FSP of the diverging train.

When determining the FSP for an approach-released signal, it must be recognised that, by imposing a more restrictive aspect, the signalling system is actually forcing the speed of the train to be reduced earlier than is required by the junction speed restriction alone. In the case of approach release from yellow, braking to the restriction can be safely achieved if it commences at (or possibly even after) the last junction protection signal. However, the display of a yellow aspect indicates that the first signal after the junction could be at red. As a result of this, in a section of 4-aspect signalling the driver would have to begin braking at the signal before this (signal 3 in Figure 8-4). For this reason, the use of approach-released signals would be expected to significantly impact capacity through a junction.

Since the FSP for a diverging train encountering a signal that is approach released from yellow will require slower speeds of operation on approach to and through the junction, the headway time will be longer than that for no-approach release.

8.3.3 DIVERGING FOLLOWED BY STRAIGHT THROUGH - APPROACH RELEASE FROM RED

Where approach release from red is used, the last junction protection signal is held at red until the train reaches a point where both the main aspect and the junction indication are readable (see section 8.3.2). Once again, the critical headway for this junction arrangement and service pattern will be that for the first junction protection signal, as represented in Figure 8-5.

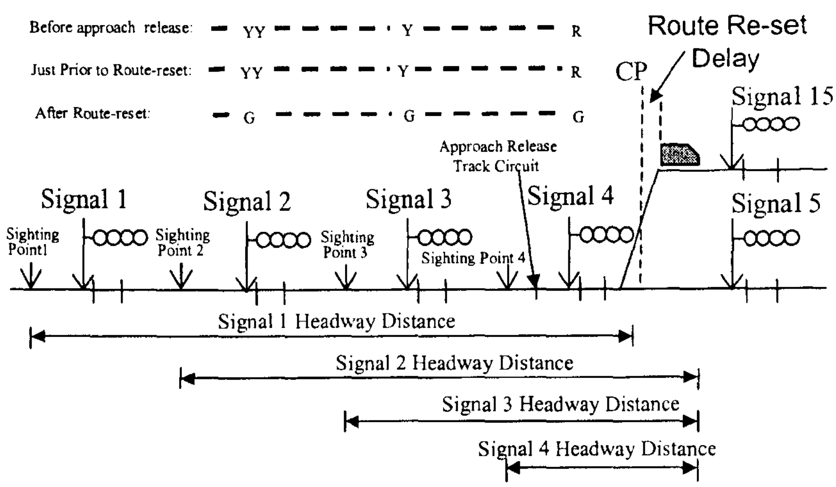


Figure 8-5: Conventional Headway for Diverging Train Followed by Straight Through Train, Approach Release from Red

As the headway distance is unchanged, the headway time for a junction that is approach released from red will still be given by equation 8-12. However, the result of this equation will again be influenced by the FSP.

In the case of approach release from red, a driver would have to brake his/her train in order to stop at the last junction protection signal (signal 4 in Figure 8-5). This requires a more restrictive approach speed than that imposed by the actual junction restriction (another example of the use of approach released signals impacting capacity through a junction).

Since the FSP for a diverging train encountering a signal that is approach released from red will require slower speeds of operation on approach to and through the junction, the headway time will be longer than that for no-approach release or approach release from yellow.

8.3.4 DIVERGING FOLLOWED BY STRAIGHT THROUGH - APPROACH RELEASE FROM FLASHING YELLOW

Approach release from flashing yellow is used where the conditions required for approach release from yellow can not be met, yet the turnout speed is higher than the train would be permitted to reach by use of approach release from red. In a three aspect signalled area, the driver of an approaching train receives a flashing yellow indication at the first junction protection signal. This acts as a route indication, advising the driver that it is the high-speed turnout route that has been set for the train. The next signal is approached displaying a yellow aspect that releases to a less restrictive aspect (conditions ahead permitting) once the train reaches a point within which both the main aspect and the junction indication are readable. In a section of 4 aspect signalling, the first junction protection signal displays a flashing double yellow aspect, which is followed by a flashing yellow and then the junction signal approach released from yellow (as shown in Figure 8-6).

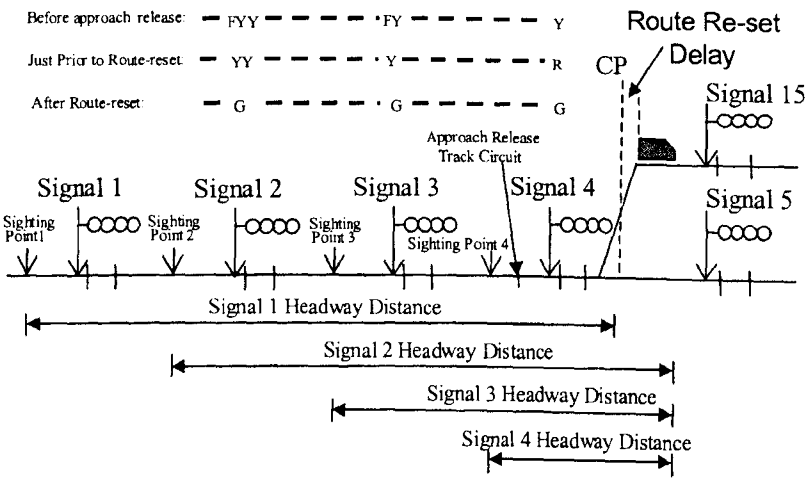


Figure 8-6: Conventional Headway for Diverging Train Followed by Straight Through Train, Approach Release from Flashing Yellow

As the headway distance is again unchanged, the headway time for a junction that is approach released from flashing yellow will still be given by equation 8-12. However, as with approach release from red, the result of this equation will again be influenced by the FSP.

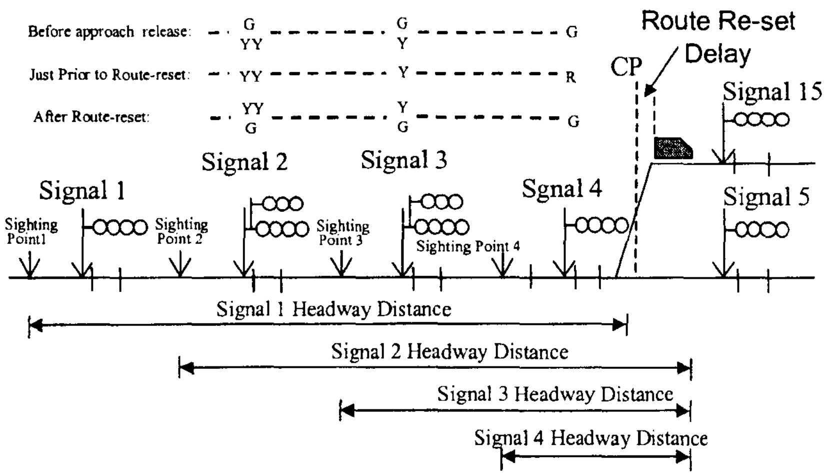
In the case of approach release from flashing yellow, a driver would have to brake his/her train in order to stop at the signal after the junction (signal 15 in Figure 8-6), due to the yellow aspect displayed by the junction signal. This requires a more restrictive approach speed than that imposed by the actual junction restriction, impacting capacity if the signal is not actually displaying a red aspect. In this case, however, since there is insufficient braking distance between the turnout and the last junction signal, the driver would need to commence braking for the junction speed before the junction signal. This means that the impact would not be so severe as that of approach release from yellow or red.

8.3.5 *DIVERGING FOLLOWED BY STRAIGHT THROUGH - SPLITTING DISTANT*

An alternative method for signalling high-speed junctions is the provision of splitting distant signals. This approach requires additional signal heads (which apply to only one diverging route), making it more expensive. It is, therefore, generally reserved for use where two consecutive junctions would otherwise require flashing aspects.

An example of the arrangement of splitting distant signals in a 4-aspect area can be seen in Figure 8-7. It should be noted that none of the splitting distant aspects are illuminated when the junction signal is displaying red (Andrews 1995, pp10-12; BRB 1992, pp13-15; Catalis Rail Training(1) 1999, pp10-11; Woolford(1) 2002, pp32-4).

The headway time is still given by equation 8-12, in accordance with the FSP applicable to the splitting distant scenario. The use of splitting distant signals allows the displayed aspect to be



based on occupancy of the line ahead, not imposing any additional restraints on the train due to approach release of signals. This means that there is no impact on the capacity of the junction when compared with provision of signalling based on no-approach release.

Figure 8-7: Conventional Headway for Diverging Train Followed by Straight Through Train, Splitting Distant Signals

8.3.6 *CONSECUTIVE DIVERGING TRAINS – NO APPROACH RELEASE*

With no approach release, the headway of consecutive trains taking the diverging route at a junction will be the same as that for a plain line section with an equivalent speed profile:

$$Ht_{n(diverging\ junction - consecutive\ diverging - no\ approach\ release)} = St + Tc + \sum_{x_{start}}^{\left(x_{start}+O+L+\sum_{p=1}^{p=n-1} Sg_p\right)} \frac{\Delta x}{V_{act}(x)} \quad \text{Equation 8-13}$$

The limiting headway for the junction will be determined by the signal for which trains experience the largest part of the restriction effect. Since a junction speed restriction is in this case no different from any other speed restriction, this means that it will be the signal before the one where braking to the restriction speed must commence that imposes the longest headway time. In the case of no approach release, the maximum differential between the diverging and straight through routes will be 5mph. Therefore, the braking distances for this scenario will be relatively low (only extending up to about 140m for a 125mph line speed). This implies that braking need not commence until after the last junction protection signal, which would be expected to have an overlap in the region of 183m. The longest headway would be that for the last junction protection signal (Signal 4 in Figure 8-8).

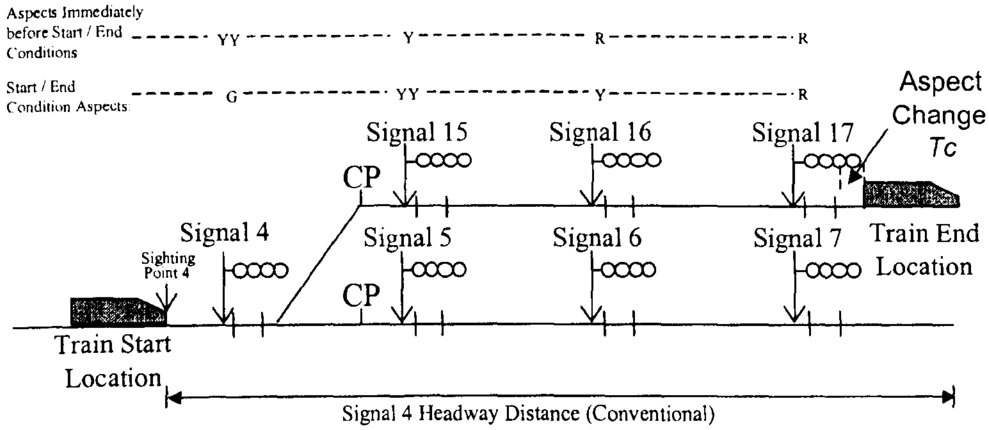


Figure 8-8: Conventional Headway for Consecutive Diverging Trains, No Approach Release

8.3.7 CONSECUTIVE DIVERGING TRAINS – APPROACH RELEASE FROM YELLOW

With the introduction of approach control, the headway distance for consecutive diverging trains differs from that of a plain line section. The critical point for the start of the headway distance is no longer the sighting point of the last junction protection signal, but its planned clearance point. This is represented in Figure 8-9. If the train arrives at this point before all other conditions required for the signal to display a green aspect are true, the signal will not clear. The displayed aspect will then restrict the train and headway will not be achieved. If the train arrives just as the other conditions required for the signal to display a green aspect become true, then it is following the leading train at minimum headway. Any later and the headway time has been unnecessarily extended.

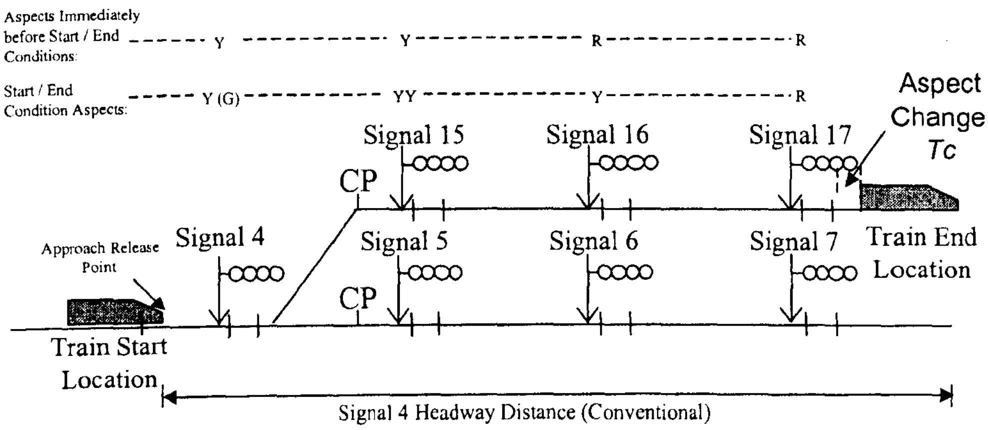


Figure 8-9: Conventional Headway for Consecutive Diverging Trains, Approach Release from Yellow

If the approach release conditions require a track circuit activated timer to delay clearing of the signal until a train is within the readable distance, it is still the intended release point, rather than the start of the track circuit that represents the start of the headway distance. If the train arrives on the track circuit before the other conditions required for clearance of the signal have been met, the timer will still start. Completion of the timer will not then cause release of the signal until the other conditions have been met.

With approach release from yellow, the headway of consecutive trains taking the diverging route at a junction will not be the same as that for a plain line section with an equivalent speed profile. It will, instead, be given by:

$$Ht_{n(diverging\ junction - consecutive\ diverging - approach\ release)} = Tc + \sum_{x_{start}}^{\left(x_{start} + S_A + O + L + \sum_{p=1}^{p=n-1} Sg_p\right)} \frac{\Delta x}{V_{act}(x)} \quad \text{Equation 8-14}$$

Where: S_A is the distance between the intended approach release point and the last junction protection signal.

8.3.8 CONSECUTIVE DIVERGING TRAINS – APPROACH RELEASE FROM RED

For consecutively diverging trains operating through a junction with approach release from red, the headway distance and equation for headway time are the same as that of approach release from yellow (see Figure 8-10 and Equation 8-14).

As with the case of a converging train being followed by a straight through train, the result of this equation will be influenced by the FSP. In the case of approach release from red, the driver would be required to follow a more restrictive speed profile on approach to a red than a yellow aspect. Since the FSP for a diverging train encountering a signal that is approach released from red requires slower speeds of operation on approach to and through the junction, the headway time will again be longer than that for no-approach release or approach release from yellow.

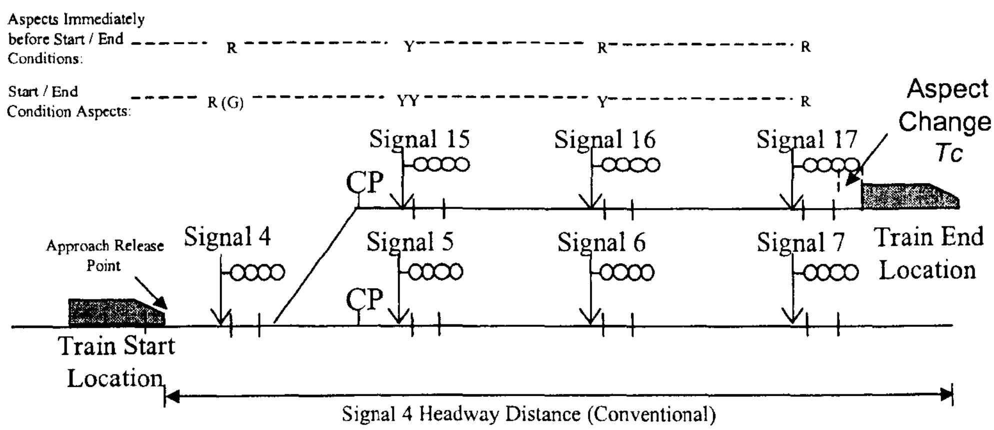


Figure 8-10: Conventional Headway for Consecutive Diverging Trains, Approach Release from Red

8.3.9 CONSECUTIVE DIVERGING TRAINS – APPROACH RELEASE FROM FLASHING YELLOW

The headway distance and equation for headway time applicable to consecutive diverging trains operating through a junction with approach release from flashing yellow will still be the same as that of approach release from yellow (see Figure 8-11 and Equation 8-14).

The result of this equation will again be influenced by the FSP. In the case of approach release from flashing yellow, a driver would have to brake the train in order to stop at the signal after the junction (signal 5 in Figure 8-11) due to the yellow aspect displayed by the junction signal. This requires a more restrictive approach speed than that imposed by the actual junction restriction, impacting capacity. However, since approach release from flashing yellow is only applied where there is insufficient braking distance between the turnout and last junction signal, the driver would need to commence braking for the junction speed before the junction signal. This means that the impact would not be so severe as that of approach release from yellow or red.

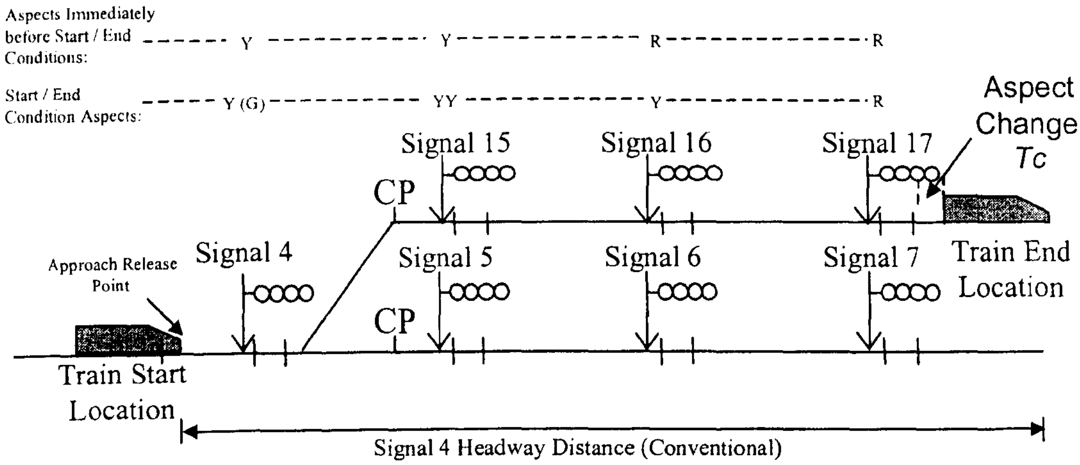


Figure 8-11: Conventional Headway for Consecutively Diverting Trains, Approach Release from Flashing Yellow

8.3.10 CONSECUTIVE DIVERGING TRAINS – SPLITTING DISTANT

In the case of splitting distant signals, trains are able to proceed without approach control and the headway of consecutive trains taking the diverging route will be the same as that for a plain line section or diverging junction with no-approach control (see Figure 8-12 and Equation 8-13). The result obtained by use of this equation will differ from no-approach release only in accordance with the FSPs applicable to the two scenarios.

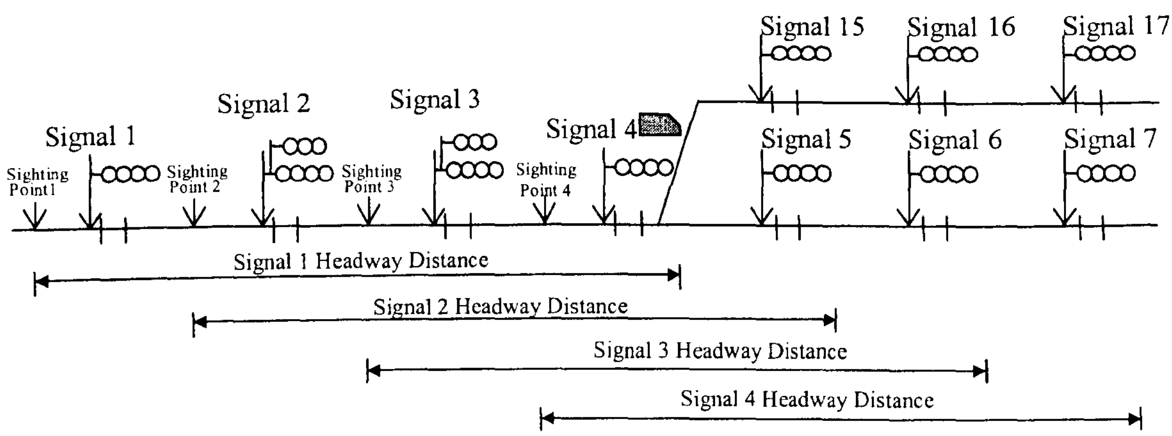


Figure 8-12: Conventional Headway for Consecutively Diverting Trains, Splitting Distant Signals

8.4 CONVERGING JUNCTIONS

Converging routes are not subjected to the approach control seen on diverging junctions. Therefore, the scenarios outlined in section 7.3.6 do not require any significant further development. However, in the light of the FSP discussions of section 8.2.1, the equation for converging junction headway outlined in section 7.3.6 can be significantly improved upon:

$$H_t = \textit{Train Following Headway} + \textit{Converging Train Run Time} - \textit{Straight Through Train Run Time} \qquad \text{Equation 7-36}$$

The converging train run time in this equation was taken to be the time from the sighting point of the first junction protection signal on the converging line (SP_{conv}) to the point at which a converging train attained the maximum permitted speed having traversed the junction (FS_{conv}).

The straight through train run time was also taken to start at the sighting point of the first junction protection signal (SP_{str} on the straight through line this time) and to end at FS_{conv}. Applying equations 8-10 and 8-11, this can now be expressed as:

$$Ht_n = St + Tc + \left(\sum_{x_{start}}^{\left(x_{start}+O+L+\sum_{p=1}^{p=n-1} Sg_p\right)} \frac{\Delta x}{V_{act(leading\,train)}(x)} \right) + \left(\sum_{SP_{conv}}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv\_train)}(x)} \right) - \left(\sum_{SP_{str}}^{(FS_{conv})} \frac{\Delta x}{V_{act(str\_train)}(x)} \right)$$

Equation 8-15

Where signal 'p' is equal to 'n' for the first junction protection signal and the 'following train' is whichever of the converging and straight through trains is last to reach its 'SP'.

In practice, the end location of the run time calculations could be any point after FS_{conv} (since both converging and a straight through trains travel with the same speed profile after that point).

8.5 MULTIPLE TRACK RAILWAYS

In terms of track infrastructure, the simplest arrangement for a railway line is that of a single track used by traffic in both directions. However, unless passing loops are added, such an arrangement limits capacity to a shuttle service. Where passing loops are provided, they can be used to allow trains in opposite directions to pass or for fast trains to overtake slow ones. The capacity is then determined by the distance between passing loops, the length of intermediate blocks, the permissible speed of the line and the service pattern (Barter 2001, p3; Schmid(1) et al. 2002, p8).

Increasing the railway to double track, with one line for each direction, provides a step change in capacity. The capacity is still limited by the permissible speed, the length of block sections and, unless passing loops or equivalent infrastructure is provided, by the speed of the slowest train travelling on the line (Barter 2001, p4; Schmid(1) et al. 2002, p8).

A further increase in capacity can be achieved by adding a third track. This is particularly beneficial where there is a 'tidal' pattern to the railway's services (that is, where services are operated to fulfil higher passenger demand in one direction than the other at certain times of the day). By stabling trains at the end of the line, ready to return along the line when the 'tide turns', the third line can be used to effectively double the tidal direction capacity. However, operation of such a service is only sustainable if the tidal flow is of a short duration, the required facilities for stabling trains are available at the ends of the line and there is sufficient rolling stock available to sustain this service pattern. "In most cases, the flow of trains in one direction must be balanced with a flow in the other direction, since trains cannot be stored in a central location to await the next traffic peak. Trains also require cleaning facilities between peak periods. Although a limited number of trains are used for peak hour services only, these are usually stored away from city centres in Britain" (Schmid(1) et al. 2002, p5).

A third line also offers significant benefits where the service pattern contains a few slow trains running amongst a larger number of fast trains, which can (with careful timetabling) be segregated out onto different lines (Barter 2001, p4; Schmid(1) et al. 2002, p9). Such an arrangement also offers potential for continuing operation of the railway at a lower capacity during maintenance of one of the lines.

The benefits of segregating traffic with different performance characteristics or stopping patterns onto different lines means that the provision of a fourth track, to achieve full double tracking in each direction, also offers capacity increases above that which would be obtained by simply doubling the two-track capacity (Barter 2001, p4; Schmid[1] et al. 2002, p9). Double tracking in each direction also provides a means of avoiding the problem of sustainability encountered with three track tidal flows.

The main problem arising from the provision of multiple tracks is the need for point work or loops to allow transfer between the lines, both at the termini and at other locations along the line. This is a particular problem (and limitation to the capacity of the line) where junctions with other lines occur.

8.5.1 JUNCTIONS ON MULTIPLE TRACK RAILWAYS

Junctions on multiple track railways may be either 'flat' or 'grade separated'. An example of typical junction arrangements on a double track line can be seen in Figure 8-13.

It can be seen that some trains travelling across a flat junction will have to cross the line used by traffic in the opposite direction. With a flat double lead junction some movements through the junction area can occur in parallel, but others must be staggered to avoid the conflict. For these trains, it is not just the preceding train in the same direction that must be clear of the junction before a green aspect can be displayed on the first junction protection signal, but also trains from the opposite direction. This results in a reduction in capacity through the junction when compared with the cases considered.

Single lead junctions, developed by BR in the 1980s in an attempt to reduce infrastructure costs, have a more severe impact on capacity – since the potential for parallel moves is reduced. In practice, the use of single lead junctions is also inherently more dangerous (with the potential for head on collisions) and less resistant to perturbations, with failures on one line immediately impacting the other (Schmid[1] et al. 2002, p10).

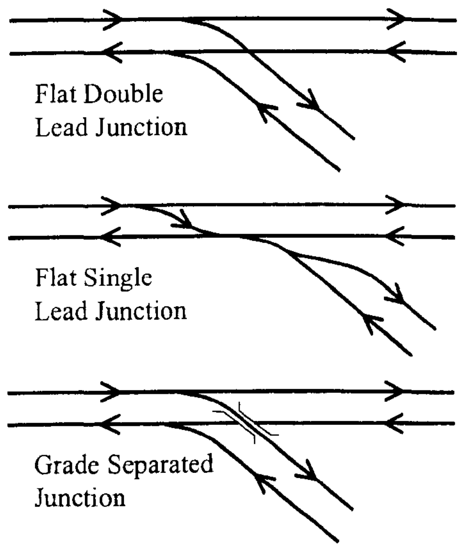


Figure 8-13: Junction Arrangements on Double Track Line

In terms of capacity, grade separated junctions are the optimal solution for multiple track railways, avoiding all conflicts between lines. They also require much less maintenance than a flat junction. However, they are also expensive to provide and require additional land usage (Barter 2001, p5).

Where the track arrangement includes flat junctions, the impact of potential conflicts must be considered when determining the practical control system capacity.

8.5.2 *MULTIPLE PLATFORMS*

Just as provision of multiple tracks will increase the capacity of plain line sections of the railway, the most effective way of increasing capacity at bottleneck stations may be the provision of parallel platforms in each direction of travel. This can allow trains to approach closer together at the stations (Gill et al 1992, p264). In order to consider the capacity of multiple platform stations, it should be borne in mind 'that any track layout, however complicated, really comprises a number of facing junctions and a number of converging junctions' (Dell 1958, p90). The use of multiple platforms requires the use of junctions between the main line and off-line platforms. Therefore, trains approaching the platform will be restricted by both the location of the train ahead and any speed restriction associated with the junctions.

On UK main line railways, a speed limit of 125mph applies to any train movements through platforms that may be occupied by passengers. On conventional LUL lines utilising trainstop protection, a speed limit of 5mph is applied on passing a station starter in order to ensure that non-stopping trains do not compromise calculated overlaps in the section ahead. Thus, a typical 100m long LUL train with a service-braking rate of 1.15m/s could approach a set of points 10m before a station platform at 36mph, whilst targeting 5mph at the starter signal (Chandler 2001, p29; Glover 2000, p48). By comparison, turnout speeds in station areas would typically be in the region of 15 to 30mph (generally being lower in confined areas, such as underground stations). A significant permitted speed differential is therefore likely between main line routes (even if there is an in-line platform) and routes to off-line platforms. As a consequence of this, the last home signal protecting the diverging junction would normally be approach released from red.

8.5.2.1 *ALL TRAINS STOPPING*

As with a single in-line platform, trains approaching a multiple platform station stop do not require green aspects to proceed unaffected by the train in front. The delay that occurs as a train enters and stops at a station also gives time for the preceding train to clear the area. As a result of this, if no approach control is applied to signals on the run into a multiple platform station signalled with 4-Aspect signalling, the first home signal (that is, three signals before the station starter) will have the most significant headway. This is also the case in 3-aspect territory with high line speeds, the first home signal then being two signals before the station starter. In both cases, the approach headway is more restrictive than that on departure from the station. However, at lower line speeds the station starter signal can be more restrictive to headway than the home signals in 3-aspect areas. A proof of this can be found in Appendix D.

When the inner home signal is approach released from red for diverging routes, the minimum headway for diverging trains approaching the station will be determined by the headway of the signal before the first home signal, that is, the last signal that can display a green aspect on approach to the station. The diverging train will then be forced to make a slower approach (due to the red aspect) than required with no approach control. This would be expected to extend the time between arrival of a straight through and diverging train. Similarly, the slow speed approach of the diverging train would extend the minimum headway between it and a following straight through train. However, the higher approach speeds possible after route reset would reduce the separation between the trains by the time they have come to rest in the platform.

For simplicity, the analysis conducted in the remainder of this section will assume that the station approach is the most restrictive to headway, that Consecutive trains enter alternate platforms and that the first train takes the diverging (approach released from red) route to the off-line platform. Based on this assumption, an example of a simple station layout incorporating one in-line and one off-line platform can be seen in Figure 8-14.

The headway distances indicated for $1^{st}/2^{nd}$ and $2^{nd}/3^{rd}$ trains respectively represent the separation required between consecutive trains in order to enable entry to the platforms without interference from the train in front. In the case of approach-released signals the signalling maintains a red aspect on the last home signal even if no train is in the route ahead. Headway is then maintained as long as the least restrictive approach aspect sequence permitted by this control can be displayed to the following train.

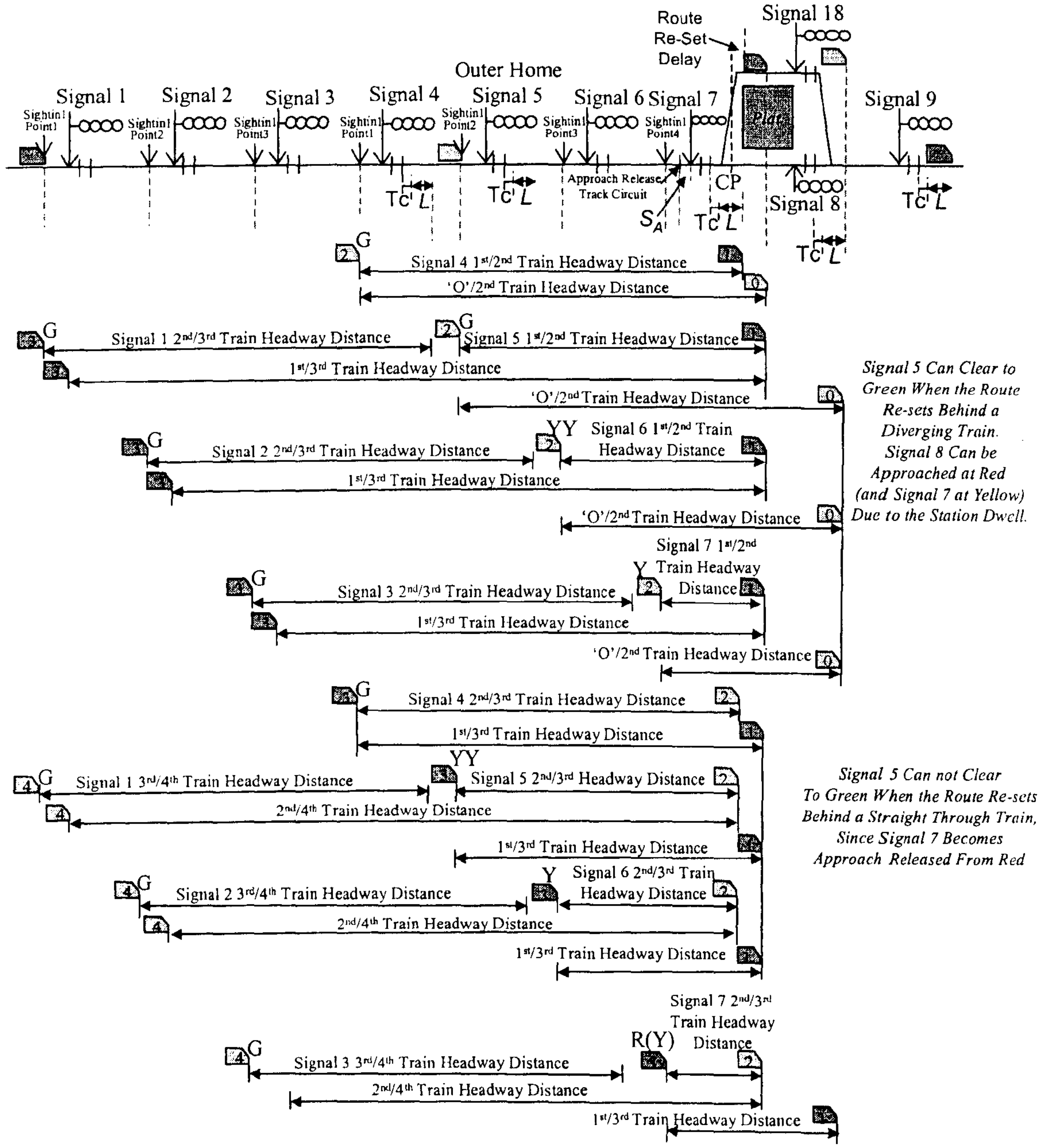


Figure 8-14: 4-Aspect Headway for Two Platforms per Direction of Travel (All Trains Stopping)

As considered in section 8.3.1, the resetting of the point lie between the 1st/2nd trains will reduce the headway distance between them on approaching the station. However, the 3rd train arriving at the station must not only remain a headway distance behind the 2nd train (which stopped in the other platform), but also the 1st train (which stopped in the same platform). For this to be achieved, the 1st train must have completed its dwell and cleared the overlap of the platform starter signal at least the signal clearance delay before the 3rd train arrives at the approach release track circuit. An equivalent situation occurs between the 2nd and 4th trains, when the 3rd train must have completed its dwell and cleared the overlap of platform starter signal at least the signal clearance delay before the 3rd train reaches the sighting point of the outer home signal. If the alternate trains arrive at the station with any smaller separation, the aspects encountered will be more restrictive than required to maintain headway. The 1st/3rd and 2nd/4th train headway distances in Figure 8-14 represent this requirement. In order to maintain headway separation between all trains, the conditions shown between consecutive trains and those shown between alternate trains must all be met.

When the headway distances are converted to headway times, the importance of considering the alternate train's headways become clearer. Whilst the duration of the station dwell will not impose a restriction on the separation of the 1st and 2nd or 2nd and 3rd trains, it will between the 1st and 3rd or 2nd and 4th trains.

The headway time required between the 1st and 2nd trains can be determined by equating the arrangement to a diverging junction (see section 8.3), with the longest headway time being that of the outer home (or first junction protection signal):

$$Ht_{n(Consecutive\ trains,\ diverging\,/\,main,\ 2\ platforms)} = St + Tc + Tr + \sum_{x_{start}}^{\left(x_{start}+Sj+L+\sum_{p=i}^{p=i+n-3} Sg_p\right)} \frac{\Delta x}{V_{act}(x)} \quad \text{Equation 8-16}$$

Where: Sj is the distance between the inner home signal and junction clearance point;

'i' is the number of the signal preceding the outer home (4 in the example of Figure 8-14).

It should be remembered that, whilst the first junction protection signal represents the longest headway through the actual junction, the plain line headway on approach must also be considered in order to determine whether the junction actually increases overall headway.

The headway time required between the 2nd and 3rd trains at the outer home will be given by:

$$Ht_{n(Consecutive\ trains,\ main\,/\,diverging,\ 2\ platforms)} = St + Tc + \sum_{x_{start}}^{\left(x_{start}+L+O+\sum_{p=i}^{p=i+n-2} Sg_p\right)} \frac{\Delta x}{V_{act}} \quad \text{Equation 8-17}$$

Since the inner home signal is approach released from red for diverging routes, the headway distance required between the 2nd and 3rd trains is smaller than that between the 1st and 2nd trains. The speed of travel over this approach released route will, however, also be lower. Therefore. the relative significance of the two headway times will be determined by the FSP of the respective routes.

As shown in Figure 8-14, on approach to the station area the headway distance required between alternate trains is determined by that required between consecutive trains. However, once the headway distance extends into the station area (where consecutive trains occupy different platforms), this is no longer the case. Whilst this results in shorter headway distances through the station area, it also means that the time associated with travel over those distances must include the station dwell. This is generally far more significant to headway time than the longer distances associated with approach headways. Therefore, for the 1st and 3rd trains (both diverging) the most significant headway will be that for the inner home signal (the first that includes the station dwell), given by:

$$Ht_{n(Alternate\ trains,diverging,multiple\ platforms)} = Tc + Dwell + \sum_{x_{start}}^{(x_{start}+S_A+Sg_{ih}+O+L)} \frac{\Delta x}{V_{act}(x)} \quad \text{Equation 8-18}$$

Where 'ih' is the number of the inner home (7 in the example of Figure 8-14).

For the 2nd and 4th trains (both straight through), the most significant headway that includes the station stop is that for the outer home signal, given by:

$$Ht_{n(Alternate\ trains,mainline,multiple\ platforms)} = St + Tc + Dwell + \sum_{x_{start}}^{\left(x_{start}+L+O+\sum_{p=oh}^{p=oh+n-2} Sg_p\right)} \frac{\Delta x}{V_{act}(x)} \quad \text{Equation 8-19}$$

Where 'oh' is the number of the outer home (5 in the example of Figure 8-14).

This is a specific instance of the general plain line headway equation derived in section 8.2.1 (equation 8-10).

Where there are two platforms provided per direction of travel, alternate train headways represent the most significant headway time due to the impact of the slower speeds of travel immediately before and after the station stop (including the dwell time). The resulting headway times would, however, be lower than those required for a single in-line platform.

Equivalent equations to those derived in this section could be developed for the headway at every signal in the junction area and for the effect of providing any number of platforms. In practice, the minimum headway time required between trains using the same platform in order to ensure that an arriving train finds a platform vacant remains as given in equations 9-18 and 9-19 for any number of platforms provided (see appendix D). However, as the number of platforms available increases, the separation imposed by the approach headway becomes more significant than the headway through individual platforms. Thus, provision of additional platforms would provide headway benefits as long as the dwell component impact on train arrivals remains longer than the interval between trains to the same platform that the plain line headway will support (i.e. as long as a train could otherwise arrive to find all platforms full). Until provision reaches this point, additional platforms reduce the impact of dwell times on following trains, enabling trains to arrive at closer intervals.

8.5.2.2 STOPPING / NON-STOPPING COMBINATIONS

Just as provision of additional platforms reduces the interdependence of consecutive stopping trains, it also reduces the interdependence of stopping/non-stopping train combinations.

The headway on approach to an off-line platform can be treated as a diverging junction with approach control from red. Thus, the minimum headway time between a stopping train in the off-line platform and a non-stopping train along the main line route will be given by equation 8-12 – with the actual headway determined by the train's speed profile on approach to the station stop. The headway for a non-stopping train followed by a stopping train is not critical, since it will be constrained by the plain line headway up to the point at which deceleration to the station stop commences, rather than the headway through the station itself.

Departure from the station can be treated as a special case of a converging junction arrangement. The equations for the headway must be revised to allow for the fact that the converging train does not need to receive a proceed aspect until completion of its dwell and for the additional station area point work. To this end, the critical headway of the whole stopping/non-stopping scenario for two platforms per direction of travel can be represented as shown in Figure 8-15.

The critical headway time for both the stopping/non-stopping and non-stopping/stopping scenarios is still given by Equation 7-36. However, when the station stop is accounted for, this gives a revised headway time equation:

$$Ht_n = St + Tc + Dwell + \left(\sum_{x_{start}}^{\left(x_{start} + O + L + \sum_{P=oh-n}^{p=oh-1} Sg_p\right)} \frac{\Delta x}{V_{act(stopping)}(x)} \right) + \left(\sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(stopping)}(x)} \right) - \sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(non-stop)}(x)}$$

for a non-stop / stopping combination, and Equation 8-20

$$Ht_n = St + Tc + Dwell + \left(\sum_{x_{start}}^{\left(x_{start} + O + L + \sum_{P=y}^{p=y+(n-1)} Sg_p\right)} \frac{\Delta x}{V_{act(non-stopping)}(x)} \right) + \left(\sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(stopping)}(x)} \right) - \sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(non-stop)}(x)}$$

for a stopping / non-stop combination Equation 8-21

Where SP is the Sighting point of the outer home signal (oh) and y is the number of the first signal after the stopping train reaches line speed (11 in the example of Figure 8-15).

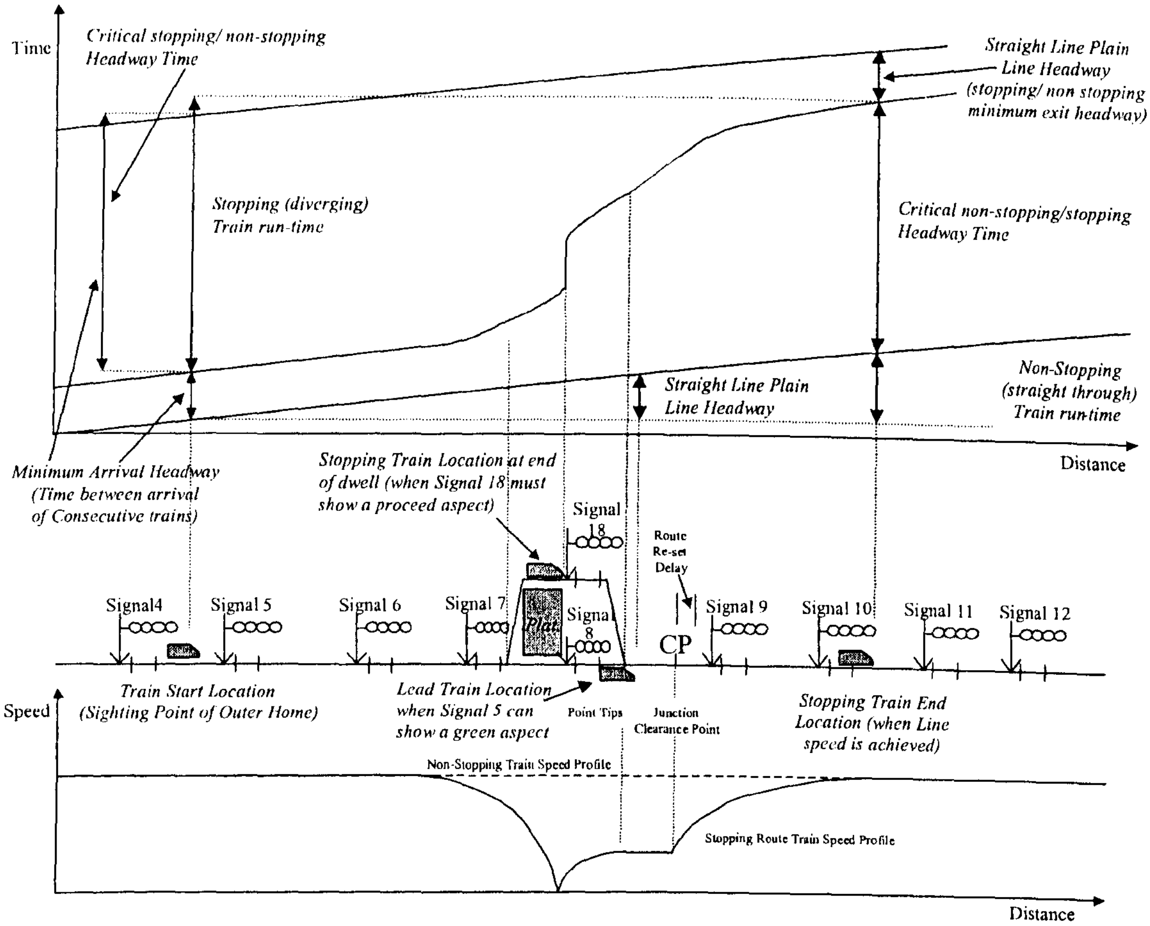


Figure 8-15: 4-Aspect Headway for Two Platforms per Direction of Travel (Alternate Stopping/Non-Stopping Trains)

As with the case of all trains stopping, the headway that can be achieved between consecutive trains is not necessarily the most significant limitation. Before a stopping train can be routed into the off-line platform, any preceding train with the same routing must have cleared the platform. If the service pattern sees alternating stopping trains (using the off-line platform) and non-stopping trains (travelling straight through on the main line), the most significant headway will be that for the inner home signal, given by Equation 8-18.

Since non-stopping trains do not dwell in the platforms, operating multiple non-stopping trains between stopping trains would increase the overall throughput of a station area with multiple platforms per direction. Not only would this allow continued use of the line during the stopping train's dwell, but it would also ensure a longer time between arrival of consecutive stopping trains that require use of the same platform, reducing the interdependence between such trains.

8.5.3 TERMINAL STATIONS

Terminal stations differ from the station arrangements considered so far in a number of ways:

- 'Passenger traffic is at its highest peak near the terminal. The stretch of line over which this peak operates is often limited in length because traffic diverging from the main line reduces the density' (Anon. 1967, p609);
- All trains passing through the station must stop;
- All platforms will have similar approach speeds, so there will be no need for approach release from red;

- Terminal station platforms often terminate in a buffer stop, with little or no margin for overrun on entry to the platform. Very low speeds of entry are therefore required to mitigate the risks associated with buffer collision, thus extending run times;
- Trains must depart in the same direction that they arrived (increasing the likelihood of conflicting movements);
- Where through station dwell times would typically be between 20 seconds and a few minutes in duration, terminal stations may typically require 7.5min for suburban / metro services, 10min for semi-fast services and 15min for fast (intercity) services (Anon. 1967, p608; Barter 2001, p9; White et al. 1998, p12). This is due to a number of causes:
  - All passengers must disembark and a new load of passengers board the train (extending the required boarding and alighting times);
  - Reversal of trains requires the shutting down of the lead cab and subsequent activation of the rear cab, which typically takes 2 to 3 minutes. The turnaround dwell can be reduced to this where it is considered cost effective to provide a second driver to board the rear cab as soon as the train stops (Barter 2001, p9). This is called 'stepping back' and is most commonly done on metro services. However, it is more usual for the existing driver to shut down the lead cab, walk to the other end of the train and then activate the rear cab;
  - As the terminal marks the end of a given journey, access may be required to clean the train and/or re-stock catering facilities.

As a consequence of these factors, 'passenger terminal handling capacity at peak times is the limiting factor which restricts suburban services more than any other' (Anon. 1967, p608). Terminal handling capacity is, therefore, critical to utilising the potential capacity of the rest of the line.

8.5.3.1 ENTERING A TERMINAL STATION

Entering a terminal station represents exactly the same scenario as travelling over a diverging junction with no approach release – simply with a lower forward speed profile. As shown in Figure 8-16, the most restrictive headway will be encountered at either the first junction protection signal (also the outer home signal) or the preceding signal, depending on the combination of forward speed profiles, junction location and route-reset delays. It is, therefore, necessary to consider both headways in order to determine the worst-case headway for trains approaching the terminal station.

The headway for the outer home signal can be determined by use of the equation already derived for the first junction protection signal of a diverging junction with no approach release (see equation 8-12 in section 8.3.1). The headway of the preceding signal (signal 23 in the example of Figure 8-16) can be found by use of equation 8-10 in section 8.2.1.

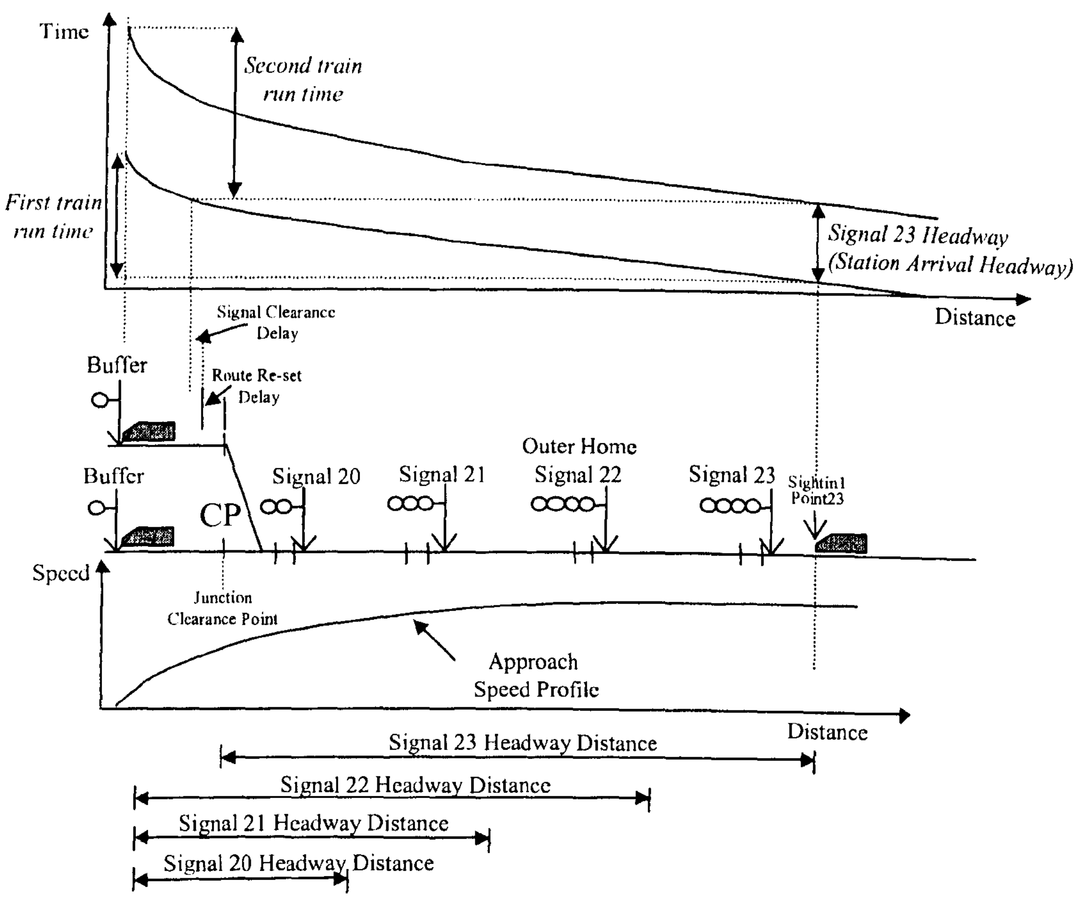


Figure 8-16: 4-Aspect Headway for Terminal Station Approach

The same equations can be used to determine the headway between consecutive arrivals at different platforms where three or more terminal platforms are fed by a single approach track. They can also be used to determine the headway leading up to a divergence where the approach track splits into two or more feeder tracks in advance of the platforms.

Some terminal stations are provided with extended overrun tracks beyond the platform ends. This reduces the risk of buffer collision, thus allowing higher approach speeds (Glover 2000, p78). The approach to such a station can be analysed in the same way, the only difference being a higher forward speed profile, which will reduce the required headway time between trains.

The author is aware of two other terminal station arrangements that would change the analysis required, because they avoid the need to reverse trains in the terminal platforms. The easiest of these to analyse is the use of a loop to connect an incoming and departing platform. All passengers arriving at the terminal must still alight at the incoming platform, but the driver does not then need to shut down his driving cab or change ends. Once the train is empty, he/she can continue around the loop to the departure platform. Outgoing passengers can then board the train ready for departure. The platform dwell time sat a terminal station arranged in this way would be expected to be longer than those of in-line stations (due to the requirement for all passengers to alight/board), but would be far shorter than a more typical terminal station. The headway associated with such an arrangement can easily be analysed by the same approach as that outlined for an in-line station in chapter 7 and for multiple platform stations in section 8.5.2.

The other potential terminal arrangement involves provision of turnaround sidings beyond the terminal platforms. This arrangement may allow higher speed approach to the station and also reduces the required dwell time. As with a loop arrangement, all passengers arriving at the terminal must alight at the incoming platform, but the driver does not then need to shut down his driving cab or change ends. Once the train is empty, he/she continues into the terminal siding. The rest of the dwell activities can then be carried out without occupying the platform. On completion of these activities, the train is driven into a departure platform, where passengers board. The use of this arrangement actually increases the overall turnaround time for the particular train being considered, but can reduce the headway impact of platform occupancy by dividing the overall dwell time between the platform and siding. The principles already outlined for headway analysis could be adapted to model this situation, but it is not proposed to analyse this scenario further in this chapter.

8.5.3.2 DEPARTING A TERMINAL STATION

When departing from a terminal station, consecutive trains pull out of different platforms and on to the same line, making it a type of converging junction. Therefore, the minimum station exit headway can be calculated in accordance with Figure 8-17 (it should be noted that whilst the figure shows the first train to have better performance than the second, this need not be the case).

In order to calculate the conventional signalling headway for this scenario, the distance taken by each train to accelerate from rest in the platform to the permitted speed of the line ahead must be calculated. The longest of these distances will define the scenario end location. The time taken by each train to run from rest in the platform to the scenario end location must then be calculated. These times are the first and second train run times.

The time that the second train can start must then be calculated by determining the time taken by the first train to run from rest in the platform to the end of signal 2's overlap, and adding any signal clearance delays (as shown in Figure 8-17). The station exit headway is then given by:

Headway time = 2^{nd} Train Start Time + 2^{nd} Train Run Time – 1^{st} Train Run Time Equation 8-22

The train run times in this equation are taken to be the time from the train being ready to start at its berth location in the terminal platform to the point at which it would attain the maximum permitted speed having traversed the junction at the station throat. Applying equations 9-10 and 9-11, this can be expressed as:

$$Ht_n = Tc + \left(\sum_{Bl}^{(Bl+Bj+Sg_1+O+L)} \frac{\Delta x}{V_{act(first\_train)}(x)} \right) + \left(\sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(second\_train)}(x)} \right) - \sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(first\_train)}(x)}$$

Equation 8-23

Where: Bl is the train berth location in the platform;

Bj is the distance between the berth location and the starter signal;

Sg1 is the distance between the starter signal and the next signal encountered.

The station exit headway must be at least equal to plain line headway. If the headway calculated in accordance with Equation 8-22 were not, the second train's start time would have to be delayed. However, starting on a yellow aspect will always ensure that trains end up separated by at least plain line headway for realistic line speeds, as shown in Appendix D.

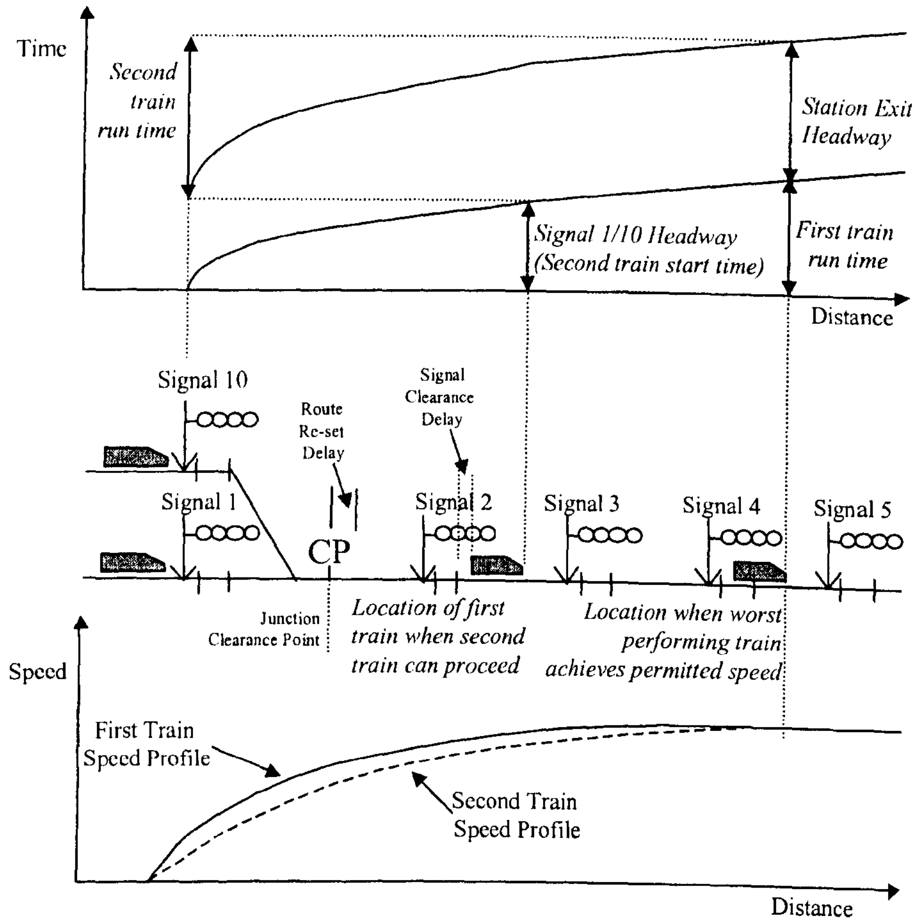


Figure 8-17: Headway for Departure from Terminal Station

8.5.3.3 OVERALL TERMINAL STATION HEADWAY

Looking at arrival and departure terminal station headways discreetly does not show the full picture of the interdependence that exists between train movements. Just as in the case of multiple platforms at a through station, an arriving train will only be able to enter its destination platform if the preceding train has already departed. In addition, since trains must enter and leave a platform at least partially over the same track, the entry and exit moves to/from adjacent platforms may conflict.

A typical terminal layout would consist of groups of four platform tracks, connected so as to permit parallel moves:

- Into platform 1 and out of platform 2;
- Into platform 2 and out of platform 3;
- Into platform 3 and out of platform 4;

The subsequent movements into platform 4 and out of platform 1 would then conflict with all other routes into or out of the group of platforms, forcing a delay between arrival of the train for platform 4 and departure of the train in platform 1, and a further delay between that departure and the next possible arrival (Barter 2001, p5).

The impact of conflicting movements will depend on the specific track layout and service patterns. In smaller stations, the length of platform dwell times makes platform occupation (dwell) the most significant factor in delaying subsequent arrivals / departures. However, in larger stations (where the number of platforms is sufficient to house all possible train arrivals for the duration of expected dwells), such conflicts can impose a large constraint on maximum capacity. Therefore, the impact of both dwell times and conflicting movements must be allowed for when determining the overall capacity of a terminal station.

8.6 PRACTICAL SYSTEM DELAY FACTORS

A number of variable factors influencing achievable headway have been highlighted within this chapter. If meaningful headway calculations are to be made possible, it is necessary to have some idea of the practical delay factors that can be expected. Therefore, the author has identified the factors and values detailed in Table 8-1.

Factor		Typical Values	Source
Gradients		±3%	Fleming 2000, p14. Also see section 6.3
Traction acceleration		$0.44 m/s^2$ below 70km/h, then decreasing linearly to $0.07 m/s^2$	Barnard(1) 2002, p31 (Based on the Class 390)
Braking performance	Service Brake	$0.88 m/s^2$ (nominal) main line, $1.15 m/s^2$ (nominal) metro	See section 5.8 for a fuller breakdown of current rates and feasible enhancements for future systems
	Emergency Brake	$1.17 m/s^2$ (nominal) main line, $1.3 m/s^2$ (nominal) metro	
Delay from route request to completion of route reset		8 seconds	TCS JPT 2002, Infrastructure work sheet
Delay from Track Circuit clearance to Signal aspect Change	For track circuit to register clearance	2 seconds (shorter delays are avoided in order to prevent track bobs registering as clearance)	Anon(3) 2002, p4; Lewis et al. 2002, pp pp28, 35-6; Appendix A Interview with Sam Macano and Jim Hoelscher
	For SSI to register change and update memory	1 to 2 major cycles, each of 608ms to 1s duration	
	For SSI to process output change	1 to 2 major cycles, each of 608ms to 1s duration	
	SSI signal module update	Typically 500ms	
	Typical overall delay for SSI	5.5 seconds	
Station Dwell	Through station	20 seconds to a few minutes. Typically 30 seconds for slow lines and 60s for fast line / intercity services	Anon(1) 2002, p41; Barter 2001, p9;
	Terminal Station	Typically 7.5min for suburban / metro services (2 to 3 min with stepping back), 10min for semi-fast services and 15min for fast (intercity) services	Anon. 1967, p608; Anon Anon(1) 2002, p41; Barter 2001, p9; White et al. 1998, p12

Table 8-1: Practical System Delay Factors

8.7 OPERATIONAL FACTORS

In addition to the system delay factors already discussed, any practical headway calculation must allow for operational factors. Variables in the performance of staff and equipment, together with externally introduced perturbations (such as extended passenger boarding/alighting times) ensure that the theoretical minimum (or technical) headway can never be achieved. In order to account for such factors, many railway administrations add a fixed 'recovery margin' to the technical headway, as already discussed in section 2.5.2. The UIC recommend that a recovery margin equivalent to one third of the technical headway be used, such that the actual operating capacity of the railway is 75% of the theoretical capacity during peak periods (Holgate 1998, p9; Holtzer 1999, p587; Schmid (1) et al. 2002, p7). In practice, a somewhat smaller margin is generally used in the UK. On metro services, a margin of 10 to 15 seconds is typically added (Gill et al. 1992, p263). For a 2-minute service interval (30tph), this would equate to between 9 and 14% of the technical headway (a 6 to 9% reduction in theoretical capacity). However, it should be noted that some additional allowance for perturbation is generally included in the dwell time assigned for each station. The technical headways used by London Underground are also calculated on the basis of speed/distance curves that allow for the practical limitations of traction, speed restrictions and other known system delays. Headway modelling for Railtrack infrastructure has also made allowances for some significant factors relating to the intended driver behaviour that can be built in to the 'technical' headway. An example of this is the allowance for normal driver braking made in train performance modelling by AEA Technology Rail, who typically assume that a driver will brake his/her train at 90% of its nominal full braking rate (Anon 2000, p1). This in itself is not considered adequate by railway authorities overseas. An example of this can be seen in the New Jersey Transit assumption of 0.522m/s^2 braking in the calculation of headway that is still de-rated by 25% for operational headway (Troup 2003, p6).

Due to the risk of signal overruns, UK operating rules and practices have changed in recent years. The 27 UK train-operating companies now all have professional driving policies that outline expected driving behavior. These policies are company specific, varying in detail between operators, but they are generally consistent with the following summary:

- At through platforms: 15 to 20mph (24 to 32km/h) 200yds (183m) before stopping point (where the AWS magnet is located);
- At terminus platforms: 10 to 15mph (16 to 24km/h) at the platform ramp, 5mph (8km/h) one or two coach lengths before the buffers and stop 2 to 6m before them;
- During permissive working: 10 to 15mph (16 to 24km/h) at the platform ramp;
- On approach to red signals: 10 to 20mph (16 to 32km/h) at the AWS magnet (200 yds / 183m before the signal). Stop 10 to 20 m before it;
- On passing a yellow aspect: Coast. Reduce speed if required for expected conditions ahead;
- On passing a double yellow aspect: reduce speed at the AWS magnet (200 yds /183m before the signal);

(Anon[(2)] 2002; Central Trains 2000, pp4-5; EWS 2001, p3; McCullie2000, pp4-5; Thames Trains 1999, pp4-5; Virgin Trains undated, pp1-2)

Some of the professional driving standards also require service braking to be limited to the use of step1 and 2 of the brake controller, which would typically equate to between 3 and 6%g (McCullie2000, pp8,21,28; Thames Trains 1999, p6; Tisi[(2)] 2000, p4).

Drivers are now trained to comply with professional driving techniques which, as a by-product of the intended increase in safety, has produced a negative impact on achievable headways. As a result, Railtrack issued a memo to timetable planners in March 2002, highlighting the need to consider professional driving practices on approach to station stops within timetable planning. This memo outlined the assumptions to be adopted in calculating section run-times, including:

- A nominal braking rate of 5%g;
- At a through platform: 20mph (32km/h) target speed achieved 200 yards (183m) before the nominal stopping point;
- At a terminal platform: 15mph (24km/h) target speed achieved at the ramp end;
- Trains stop 20 yards (18m) before the nominal stopping point.

(Appleby 2002, p1)

It is unclear whether this statement was intended to apply only to station stops, or to all stops (such as braking to a red signal). However, since professional driving approaches have been introduced specifically to prevent SPADs and most state that they apply equally to station stops and red signal approaches, it seems sensible to assume that cautious braking on approach to a station stop would be replicated on approach to a red signal (a factor of some significance to headway calculation where signals are approach controlled). However, there is no suggestion that the assumptions outlined should also apply to braking to a speed restriction.

On New Jersey Transit, a similar braking rate of $0.522 m/s^2$ (5.32%g) is assumed in headway modelling (Troup 2003, p4).

In light of the information available on expected driver behaviour, it is the author's intention to adopt the characteristics shown in Table 8-2 to represent current driving practice in the forward speed profiles used to calculate headway times.

Activity		Expected Behaviour
Station Stop	At through platform	Nominal braking rate of 5%g, to achieve 20mph (32km/h) target speed 200 yards (183m) before the nominal stopping point. Then coast until braking is required at a nominal rate of 5%g to stop 18m before the stopping point.
	At terminus platform	Nominal braking rate of 5%g, to achieve 15mph (24km/h) target speed at the ramp end, 5mph (8km/h) one coach length before the buffers and stop 5m before them. After each target point coast until braking is required at a nominal rate of 5%g for the following target.
	During permissive working	Nominal braking rate of 5%g, to achieve a maximum of 10 mph at the platform ramp.
On approach to red signals		Nominal braking rate of 5%g, to achieve 20mph at the AWS magnet (200 yds / 183m before the signal). Then coast until braking is required at a nominal rate of 5%g to stop 18m before it
Passing a Cautionary Aspect	Yellow aspect	Coast even if speed reduction not required yet. Subsequently reduce speed when required for expected conditions ahead, with a nominal braking rate of 5%g
	Double yellow aspect	Reduce speed at the AWS magnet (200 yds /183m before the signal), with a nominal braking rate of 5%g, coasting once train speed reaches 20mph (until AWS magnet of next signal is passed)
Approaching a Speed Restriction		90% of trains nominal full service braking rate

Table 8-2: Current Driving Practice

8.8 SUMMARY

Within this chapter, the author has derived equations for practical headway, allowing for actual signal spacing signalling system delays (including both processing delays and designed restrictions, such as approach release of signals). The derived equations also allow for inclusion of train performance characteristics and operator behaviour, through the forward speed profile.

Equations have been developed for a variety of signalling scenarios, including junctions, through stations and termini. Whilst these have been limited to fixed block multiple aspect signalling with a nominal overlap, the same principles could be expanded to encompass alternative arrangements, such as block section overlaps and traditional LUL signalling.

Typical values for the delay factors imposed by the signalling system, track, train performance and operator behaviour have also been identified, in order to provide the information required to develop the derived equations into practical headway models.

Whilst the equations included within this chapter provide a basis for modelling conventional UK main line signalling systems, they do not consider more advanced signalling including ATP overlay systems, in-cab signalling or moving block, which will be considered in chapter 9.

9 ATP TRAIN CONTROL SYSTEM CAPACITY

9.1 INTRODUCTION

In this chapter the equations for practical fixed block UK main line signalling headways (derived in chapter 8) will be modified and developed further to represent the headway of operation with ATP overlay and in-cab signalling systems. In order to do this, five ATP system arrangements will be considered:

1. An intermittent ATP overlay system utilising spot transmission points (beacons) or loops to transmit supervision data to passing trains. This arrangement is typical of the two intermittent ATP systems currently in use in the UK (the Chiltern and Great Western Line BR-ATP trial projects) as well as the proposed ERTMS level 1 systems described by the ERTMS Programme Team as system 'A' and 'B' (depending on in-fill provision);
2. A continuous ATP overlay system utilising radio based data transmission. This arrangement would be typical of the ERTMS level 2 overlay system previously proposed for implementation on the West Coast Main Line and described by the ERTMS Programme Team as system 'C';
3. A continuous fixed block ATP in-cab signalling system. This arrangement would be typical of the ERTMS level 2 configuration without lineside signals currently being proposed for implementation in the UK, described by the ERTMS Programme Team as system 'D'. Alternatively, ERTMS Level 3 implemented in a fixed block arrangement (mapping reported train locations onto data defined fixed blocks prior to determining movement authorities), which was not considered by the ERTMS Programme Team, would also represent this arrangement;
4. A continuous moving block ATP in-cab signalling system. This arrangement would be typical of the ERTMS level 3 configuration, which was not considered by the ERTMS Programme Team;
5. A continuous moving block ATP in-cab signalling system operating with relative, rather than full, braking distance separations. The ERTMS Programme Team did not consider this arrangement.

   (ERTMS Programme Team 2002, p9)

The derivation of equations to represent each of the ATP system arrangements for all scenarios considered in chapter 8 is repetitive, but none the less necessary to identify the differences between them. In order to assist the reader, derivations are only included in this chapter for plain line headways. If the reader wishes to see the full working and explanation of differences between each ATP arrangement for the remaining scenarios, these can be found in Appendix F.

9.2 INTERMITTENT ATP OVERLAY

Any overlay ATP system must operate in conjunction with the signalling system underlying it. The driver of an equipped train must continue to comply with the instructions transmitted through the lineside signal aspects, whilst also pre-empting intervention by the ATP system. In an ideal application, complying with the intended response to signal aspects would be adequate to ensure avoidance of ATP intervention. In such a case, the ATP system would not impact on the conventional signalling headway. However, practical applications of intermittent ATP systems (including both of the BR-ATP systems and ERTMS level 1) rely upon the state of the underlying signalling system to determine the supervision criteria enforced by the ATP. This imposes a delay between changes in the state of the underlying signalling system and subsequent update of ATP transmission points (balises or loops). In addition to this, because intermittent transmission means that trains can only receive updates to their supervision criteria at defined locations, delays can occur between updated supervision criteria becoming available at a balise or loop and the subsequent arrival of an approaching train. Therefore, the location of update points is critical to the achieved headway.

The optimal balise location on approach to a signal will be determined by a combination of the train's approach speed and the ATP system's supervision algorithms and response times. This means that a balise location can only be optimal for one approach speed profile. A train arriving at a higher speed would require an update of supervision criteria at an earlier point, whilst a lower speed train would not need the update so soon. This highlights a dis-benefit of intermittent ATP, in that it imposes additional delays under degraded operating conditions. It also highlights the need for provision of in-fill balises in order to minimise such delays and enable operation of stock with different performance characteristics as close to optimally as is practicable.

Before moving on to the derivation of equations for intermittent ATP overlay headway, one more factor needs to be considered. Since headway is achieved when the train ahead does not affect a following train, the minimum headway time permitted by an ATP system will be influenced by its supervision criteria on approach to a signal. As with lineside signalling, to maintain headway the following train must not receive an indication of the need to brake.

Assuming the ERTMS typical braking curve supervision characteristics outlined in Figure 6-2 to be typical of any ATP system, this means that the effective 'sighting point' under ATP operation (when indication of the need to brake is given) will be a time before the service brake intervention point given by:

- The time allowed between indication of the need to brake on approach to the limit of a movement authority and the actual permitted speed being crossed if the train speed is not reduced (T_{ip})
- The time allowed between crossing the permitted speed curve on approach to the limit of a movement authority and issuing of a warning if the train speed is not reduced (T_p) and;
- The time allowed between warning and service brake intervention on approach to the limit of a movement authority, if the train speed is not reduced (T_w).

The service brake intervention point will in turn be dependent on the braking rate assumed within the ATP supervision algorithms, which would typically allow for a delay in traction cut off, commencement of brake and subsequent brake build up.

For the purposes of generic headway comparison, the author proposes to use a simplified braking algorithm for the equations to be derived in this chapter. To this end, allowance will be made for a nominal delay in brake build up (bbu), equivalent to the brake application delay and half of the brake build up time. Allowance for traction cut off delays will not be made.

In light of these factors and assumptions, the optimal balise location can be determined by calculating the distance required between a signal and its 'sighting' balise if headway is to be maintained. This distance will be given by the sum of the braking distance, brake build up, margins and processing time required to ensure that the ATP system could still stop the train before the signal if the balise message were found to indicate that the signal was displaying a red aspect. If there is not a balise in the optimal location, the 'sighting' balise will be the preceding balise, an 'update location error' (Ule) before the optimal location. A representation of this is shown for signal 4 in Figure 9-1.

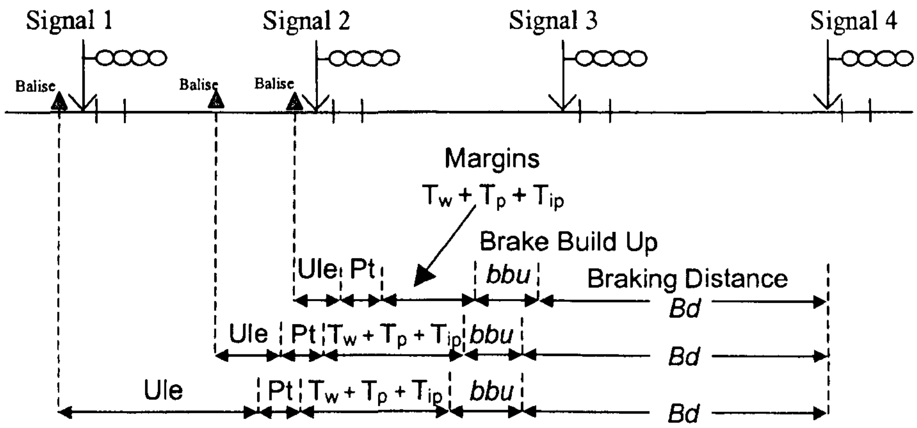


Figure 9-1: Intermittent ATP 'Sighting' Balise Location

Depending on the ATP system being modelled, the braking rate assumed on completion of build up may then be a nominal rate, or speed dependent values based on expected performance of the particular train type's brakes. The assumed braking rate could also make allowance for areas of poor adhesion where the expected braking rates would not be achieved, equivalent to the allowance of 10% (subject to a minimum amount of 10m) added during calculation of stopping distance for conventional signalling (Fleming 2000, p25).

Accounting for the balise location(s), the headway of an intermittent overlay ATP system using balise based update transmissions can be represented as shown in Figure 9-2.

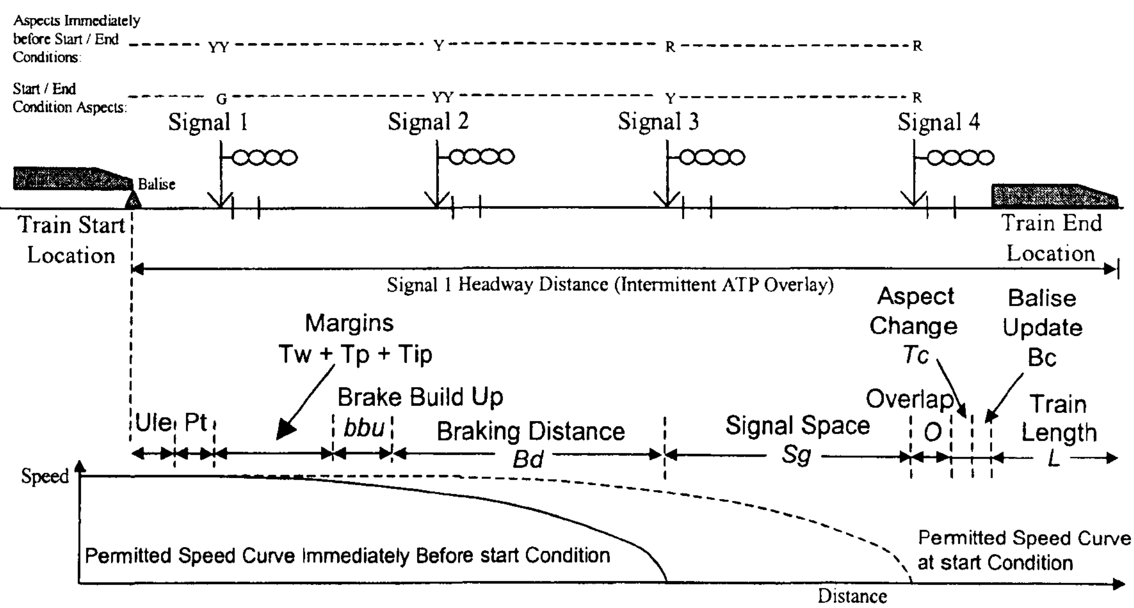


Figure 9-2: Intermittent ATP Overlay Headway

Applying the principles for calculating headway time outlined in chapter 8, the headway of a balise based intermittent ATP overlay system is, therefore, given by:

$$H_{IATP-O} = \sum_{x_{start}}^{(x_{start}+Ule+Pt+Tw+Tp+Tip+bbu+Bd+Sg+O+Tc+Bc+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation 9-1}$$

Where:

B_c is the time required for the message that will be transmitted by a balise to change, once the conditions for it to do so have been satisfied;

P_t is the trainborne ATP system's processing delay between receipt of the balise message onboard and availability of the new supervision criteria to the trainborne systems;

Bd is the braking distance, calculated on the basis of the TCS service brake curve.

Pt, Tw, Tip, bbu, Tc, Bc and Ule, are delays (and therefore times) used in calculating the optimal ATP balise location. The location is determined by converting the times to distances, based on the assumed speed of the train before they would commence. It is these distances, rather than the original times, which are implied by their use within the headway time equation. As can be seen from the equation, these distances are then divided by $V_{act}(x)$, which converts them back into times. The resulting headway component times will only be equal to the original delay times where $V_{act}(x)$ is equal to the originally assumed speed of the train throughout the associated distance.

In order to mitigate some of the headway impact that can be experienced due to sub-optimal balise location and to improve system response to degraded or perturbed operating conditions, track loops can be used in place of balises. Whilst the area over which trains can receive transmissions from balises is limited to, at most, a few metres, loops make it possible to receive transmissions over a much larger area – dependent on the size of loop used. This makes it much more likely that a train will be able to receive transmissions at the optimal location, reducing the impact of location errors, as represented in Figure 9-3.

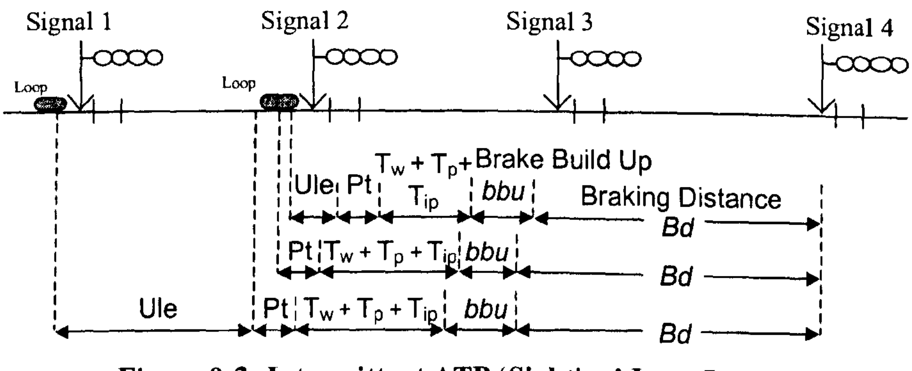


Figure 9-3: Intermittent ATP 'Sighting' Loop Location

If the ATP 'sighting' balise or loop is at or before the sighting point of the underlying signalling, the ATP will impose an increase in headway distance (which will in turn result in an increase in headway time). This may also be the case where the sighting point of the underlying signalling is slightly before the ATP sighting balise/loop (due to the balise/loop update delay). However, if the ATP sighting balise/loop is only required to be well within the sighting point of the underlying signalling, the ATP may not act as a constraint on achievable headway. In that case, the headway will be determined by applying equation 8-10 rather than Equation 9-1.

When calculating the headway applicable with an ATP system, the forward speed profile used to calculate $V_{act}(x)$ must be in accordance with the most restrictive of the ATP system's permitted speed curve and the driver's brake curve at any given point during the train's journey through the route being considered. That is, it must represent the speed profile that the driver would have to follow in order to comply with the lineside signal aspects, pre-empt intervention by the ATP system and comply with professional driving practice. In the case of a plain line section with trains operating at headway separation, the only restrictions to the trains' progress to be considered would be speed restrictions, on approach to which compliance with professional driving practice for lineside signalling would require the driver to brake at a rate equal to 90% of the train's nominal full service braking rate (see discussion of current driving practice, section 8.7). In comparison, the calculation of most ATP system's permitted speed curves would be based on the full service-braking rate. Assuming accurate targeting of speed restrictions within the ATP system data, even with an assumed increase in braking distance of 10% (equivalent to a reduction in brake rate to 91% of the train's nominal full service braking rate), this would always make the driver's brake curve the most restrictive speed profile.

9.3 CONTINUOUS ATP OVERLAY

It should be noted that there are three possible supervision criteria update strategies under a continuous ATP arrangement. Updates can be event driven (such as on occupation or clearance of a track circuit), request driven (when the trainborne system must requests an update before it is sent) or time based (say, every 1 second). If event driven, the delay occurring between the availability of updated supervision criteria trackside and on the train is the message transmission time. If request driven, an additional delay may occur between message availability trackside and the receipt of a transmission request from the train it applies to. This delay would have the effect of increasing the headway time, but may be implemented in order to reduce the number of messages that must be sent to each train and, thus, the volume of radio traffic required within the system. In some respects, time interval updates are a form of event driven update – the event being expiry of the next time cycle. However, a delay of up to one time interval could potentially be introduced between the conditions for a movement authority extension being met and the next transmission time slot. This would again increase the headway time, although the headway impact would be small so long as the time interval is kept small.

In the subsequent development of headway equations, the author will assume that supervision criteria updates are event driven. In the case of systems including train based position location, one of the events would be expiry of the next reporting time cycle, with additional events (such as route confirmation) also acting as movement authority update triggers.

As with the case of intermittent ATP systems, practical continuous overlay systems (such as ERTMS level 2) rely upon the state of the underlying signalling system to determine the enforced supervision criteria. This means that a delay will still be imposed between changes in the state of the underlying signalling system and subsequent update of ATP transmissions. However, the use of continuous transmission means that trains can receive updates to their supervision criteria at any location wgich, with event driven updates, minimises update delays.

Adopting the supervision algorithms already outlined in section 9.2, the headway of a continuous overlay ATP system can be represented as shown in Figure 9-4.

Applying the principles for calculating headway time outlined in chapter 9, the headway of an intermittent ATP overlay system is given by:

$$H_{CATP-O} = \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+Sg+O+Tc+Ut+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation 9-2}$$

Where U_t is the ATP system transmission delay, including both time for the updated message to become available once the conditions for it to do so have been satisfied and the actual transmission time.

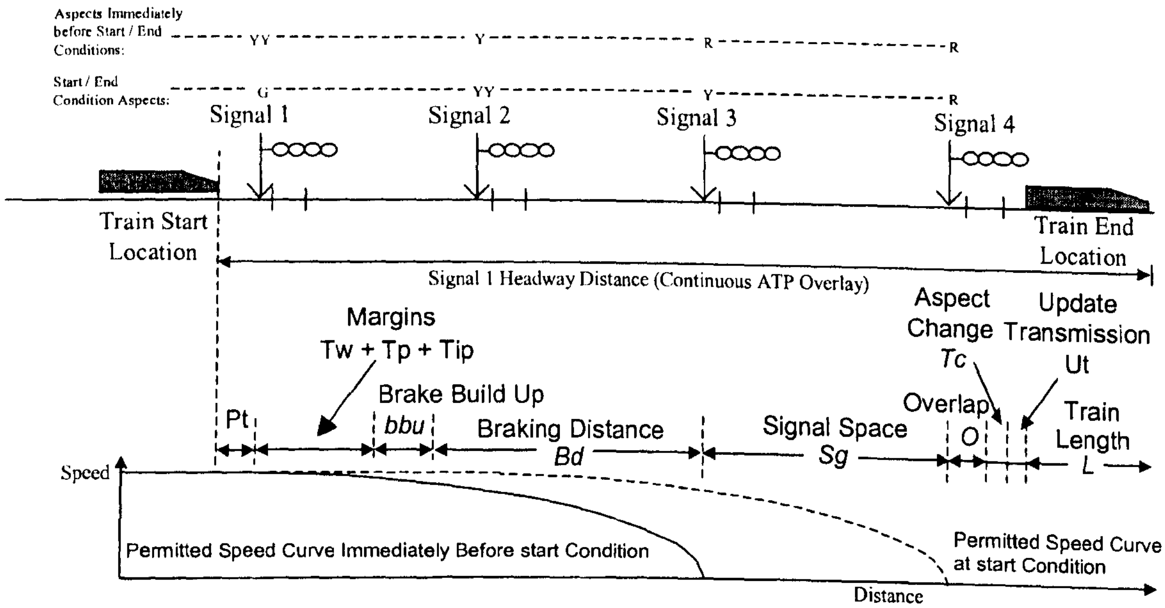


Figure 9-4: Continuous ATP Overlay Headway

As in the case of intermittent updating of supervision criteria, if the ATP 'sighting' point (shown as the train start location in Figure 9-4) is at or before the sighting point of the underlying signalling, the ATP will impose an increase in headway distance (which will in turn result in an increase in headway time). This may also be the case where the sighting point of the underlying signalling is slightly before the ATP sighting point (due to the update transmission delay). However, if the ATP sighting point is only required to be well within the sighting point of the underlying signalling, the ATP may not act as a constraint on achievable headway. In that case, the headway will be determined by applying equation 8-10 rather than Equation 9-2.

9.4 CONTINUOUS IN-CAB ATP

With the provision of continuous in-cab displays, lineside signals are no longer required. This means that the driver is free to drive solely in accordance with the in-cab display. It also means that update need not be constrained by signal locations, but can be associated with any train detection section. In consequence, a reduced headway distance can be achieved with the same number of track sections as would be required for lineside signalling. This is demonstrated in Figure 9-5. Where the system is required to overlay a fixed signalling system as a part of a dual signalled area (permitting operation of unfitted trains between fitted trains), the driver would be required to observe both lineside and in-cab signals. The author considers this situation to be a case of continuous ATP overlay, as already discussed in section 9.3. Therefore, it will not be considered within the discussion of continuous in-cab ATP.

The value Tdc in Figure 9-5 is the time taken for a change in state of the train detection system to be registered once the conditions for it to do so have been physically satisfied. This would be less than the time taken for a conventional signal aspect to change (Tc), which includes Tdc.

The headway examples of Figure 9-5 show that it is the combined distance of two consecutive train detection sections that determines the headway distance. Making the sections equi-distant or consistently equivalent to combinations of conventional signal sections and overlaps does not, therefore, change the headway distance (it simple determines the regularity of supervision updates). However, if one or more of the track sections is reduced in length without increasing the length of its neighbouring sections, the headway distance over that section reduces. Whilst this would make no difference in practical terms to the line headway of a plain line section with a uniform speed profile (since the surrounding area would still have the longer headway distance), it could be used to reduce the impact of speed restrictions or station stops (effectively allowing the headway distance to be varied with speed in order to maintain a consistent headway time throughout the line). In accordance with this approach, a more optimal headway can be achieved by extending the length of train detection sections in non-critical areas and applying numerous short sections in critical areas (such as stations). Since an in-cab ATP system does not require the concept of repeater signals, sections need not be constrained to the same minimum distances required in lineside signalling (braking distance for 3-aspect and an average of half braking distance for 4-aspect sections). This represents a significant advantage of continuous in-cab ATP systems.

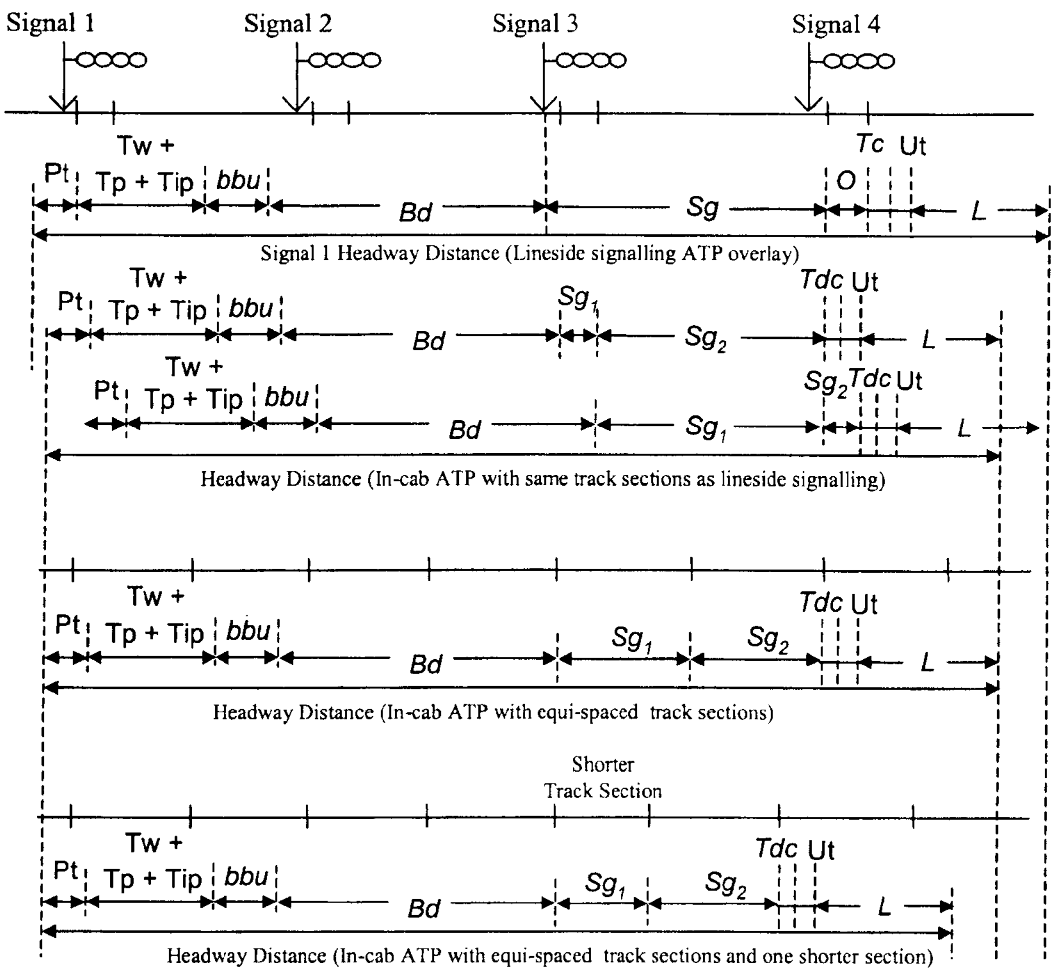


Figure 9-5: Fixed block Headway Comparison of Overlay and In-Cab ATP

Applying the principles for calculating headway time outlined in chapter 8 to Figure 9-5, the headway of a continuous ATP overlay system is given by:

$$H_{CATP-In} = \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+Sg_1+Sg_2+Tdc+Ut+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation 9-3}$$

This equation would apply equally to a system based on track circuits, axle counters or any other method of train detection (including train based location reporting, as used by ERTMS level 3), so long as the implementation of the system continues to require block sections of fixed lengths to become clear before updating supervision targets and the lengths of those track sections are at least the equivalent of an overlap length.

9.5 MOVING BLOCK

With the introduction of moving block, the location of each train is generally determined on-board, based on a combination of speed measurement sources (such as tachometers, accelerometers and doppler radars), coupled with absolute position locators (generally in the form of trackside balises). The exception to this is the approach adopted by GE's AATC system, where trains determine their distance from multiple fixed trackside radio units, based on comparison of transmission and receipt times of periodic radio signals, and send the distances back to a central processor which then calculates the train's location by use of triangulation (see Appendix C). In either arrangement, "physics dictates that the mere act of determining position must occupy a finite time – therefore if the vehicle is moving, once its position is determined, it is no longer there" (Riley 1999, p2). As represented in Figure 9-5, the positional error resulting from the latency of location information must be accounted for in determining headways.

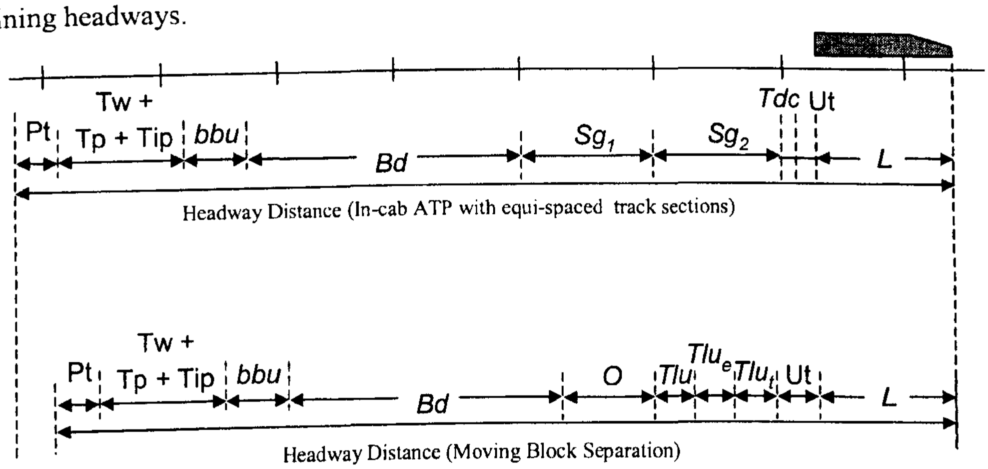


Figure 9-6: Moving block Headway, Compared to Fixed Block In-Cab ATP

Applying the principles for calculating headway time outlined in chapter 8 to Figure 9-5, the headway of a moving block system is given by:

$$H_M = \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+O+Tlu+Tlu_e+Tlu_t+Ut+L)} \frac{\Delta x}{V_{act}} \qquad \text{Equation 9-4}$$

Where: O represents a safety margin, equivalent to the overlap used in fixed block signalling;

Tlu represents the train location update interval (the time duration between the issuing of a location update by an individual train);

Tlu_e represents the error that is introduced into a train's estimate of its own location, due to factors such as slip / slide, between absolute position location devices;

Tlu_t represents the train location update transmission delay. That is, the time delay between capture of location measurement data and receipt of a location message by the central processing system;

U_t represents the ATP system transmission delay, including both time for the updated message to become available once an updated train location has been received or other conditions required have been satisfied and the actual transmission time.

It should be noted that the value 'Tlu_e' would in practice vary as the train progresses through a section of line, being reset to zero whenever an absolute location marker is passed. The error may be positive or negative. In the former case (where a train appears to be further along the line than it actually is), the impact of the error is to allow trains closer together than should be permitted, effectively reducing the safety margin available. In the latter case, (where a train appears not to be as far along the line as it actually is), the impact is to reduce the extent of the following train's movement authority, effectively extending the headway distance. When considering safety it is, therefore, a positive error that must be assumed, whilst when analysing headway, the negative error must be incorporated.

9.6 RELATIVE BRAKING

The concept of relative braking separation is discussed in some detail in Appendix F. It is based on the fact that a train cannot stop instantaneously. Therefore, if the speed and braking performance of the lead train are known, a movement authority can be given to the following train based on the earliest stopping point of the lead train, rather than its last reported location.

The headway distances associated with relative braking operation can be seen in Figure 9-7.

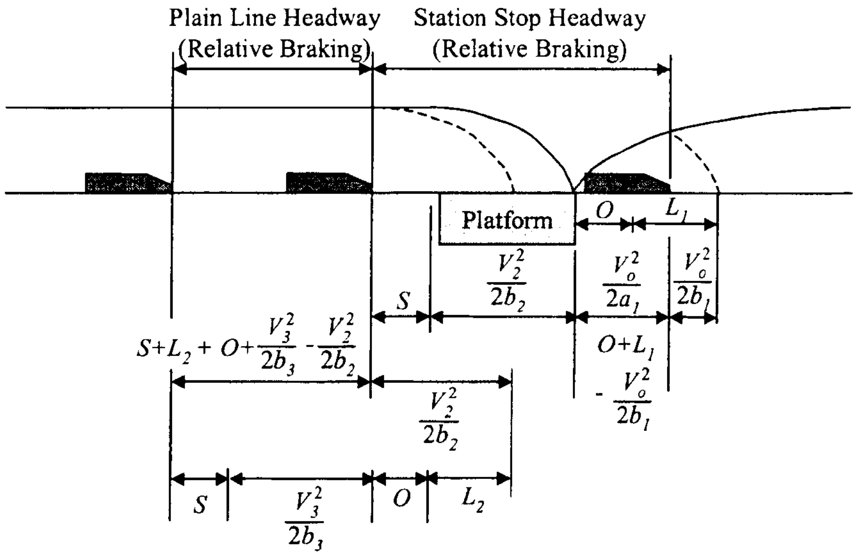


Figure 9-7: Relative Braking Headways

The plain line headway time under relative braking operation is, therefore, given by:

$$H_{t\,rb} = \left(\frac{V_f^2}{V_l \times 2b_f} - \frac{V_l}{2b_l} \right) + \frac{S + O + L_l}{V_l} \qquad \text{Equation 9-5}$$

Where V_l and V_f are the maximum permitted speeds for the lead and following trains respectively.

The headway distance is converted into time by use of the lead train speed, so that the result represents the time that will elapse between occupation of the start location by the lead train and the following train reaching the same location whilst remaining headway distance behind the lead train. That is, the headway time interval between the trains passing the start location.

Following a similar approach for the headway time between trains stopping in a platform gives:

$$H_{t\,rb(Stopping)} = \frac{S + \frac{V_f^2}{2b_f} - \frac{V_l^2}{2b_l}}{V_l} + \frac{V_l}{b_l} + Dwell + \begin{cases} \sqrt{\dfrac{2\left(O + L_l - \frac{V_o^2}{2b_l}\right)}{a_l}} & \text{For } V_l \geq \sqrt{2a_l\left(O + L_l - \frac{V_o^2}{2b_l}\right)} \\ \dfrac{V_l}{2a_l} + \dfrac{O + L_l - \frac{V_o^2}{2b_l}}{V_l} & \text{For } V_l \leq \sqrt{2a_l\left(O + L_l - \frac{V_o^2}{2b_l}\right)} \end{cases}$$

Equation 9-6

Where V_o is the speed at which the lead train should be travelling for optimum headway on departure from the platform as the following train approached the platform.

V_o can be determined by finding the speed for which the leading train's acceleration distance and braking distance from / to rest would be equal to the train length plus the safety margin:

$$V_o = \sqrt{\frac{O+L}{\frac{1}{2a_l}+\frac{1}{2b_l}}} \quad \text{for } V_l \rangle \sqrt{\frac{O+L}{\frac{1}{2a_l}+\frac{1}{2b_l}}} \quad \text{and otherwise } V_o = V_l \qquad \text{Equation 9-7}$$

If this theoretical scenario is developed to allow for practical headways, as outlined in chapter 8, the headway becomes as shown in Figure 9-8.

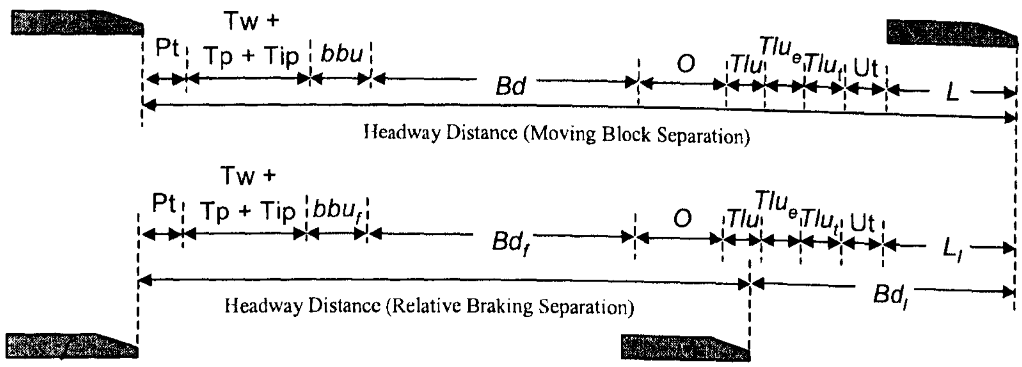


Figure 9-8: Relative Braking Headway, Compared to Moving block Headway

The plain line headway time is then given by:

$$H_{t\,rb} = \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu_f+Bd_f+O+Tlu+Tlu_e+Tlu_l+Ut+L_l-Bd_l)} \frac{\Delta x}{V_{l\,act}} \qquad \text{Equation 9-8}$$

9.7 ATP SYSTEM DELAY FACTORS

A number of variable factors specifically related to ATP operation have been identified in this chapter. Typical values for these factors are given in Table 9-2 and Table 9-2.

Factor	Typical Values	Sources
Trainborne ATP system processing delay (Pt)	1s	TCS JPT 2002, System work sheet; Lewis et al. 2002, pp27 & 29; Appendix A interviews with Sam Macano, Jim Hoelscher and Mark Glover
Margin between indication of the need to brake and crossing the permitted speed curve (Tip)	4s	TCS JPT 2002, System work sheet
Margin between warning and service brake intervention (Tw)	6s	
Time for the message that will be transmitted by a balise / loop to change once the conditions for it to do so have been satisfied (B_c)	2.4s (including time for train to read balise)	TCS JPT 2002, System work sheet; Lewis et al. 2002, pp27 & 29
Margin between crossing the permitted speed curve and issuing of a warning (Tp)	3s	ERTMS Users Group(1) 2002, p11
Continuous ATP system transmission delay (time to update and transmit message trackside) (U_t)	5s (including 3s processing t and 2s transmission)	TCS JPT 2002, System work sheet; Lewis et al. 2002, p27; Appendix A interviews with Sam Macano, Jim Hoelscher and Mark Glover
Brake build up for service brake (bbu)	3.5s	Anon(5) 2002 (First Great Western class 43); Lewis et al. 2002, p36
Brake build up for emergency brake (not used in headway calculation)	2s	

Table 9-1: ATP System Delay Factors

Factor		Typical Values	Sources
Train location update interval (Tlu)	Traditional train based speed/location measurement via tachometers, etc.	1s	SACEM (Singapore MRT) 0.624s (Appendix A interview with Michel Carnot); ALCATEL (London DLR) 1s (Uebel 1998, p6; Appendix A interview with Andrew Dalgleish); Sao Paulo Metro system 1s (Demetrescu et al. 1995, p89)
	Radio range finding	0.5s	AATC (San Francisco BART) 0.5s (Anon[(1)] 1999, p5; Anon[(2)] 1999, p19; Appendix A interview with Mervyn Parvard)
Train location update error (Tlu_e)	Traditional train based speed/location measurement via tachometers, accelerometers, etc.	20m (max.)	The ERTMS based TCS project assumed +/- 1% of distance travelled (Lewis 2002, p29); The TASS project assumed $\pm$10% for a class 390 and $\pm$3m $\pm$1.5% for a class 221(Barnard[(1)] 2002, pp36 & 39). Assuming 1km between absolute position markers, accumulated error could reach 10m, 18m and 100m respectively.
	Radio range finding	5m (max.)	AATC (San Francisco BART) determines train location to within 15 feet (4.6m) at any location (Anon[(1)] 1999, p1)
Train location update transmission delay (Tlu_t)		3s	TCS JPT 2002, System work sheet; Appendix A interviews with S Macano, J Hoelscher and MGlover
Safety margin (O)	Traditional train based speed/location measurement via tachometers, accelerometers, etc.	25m up to 105km/h (65mph) 100m at higher speeds	The Westinghoue system for JLE proposed an 80m safety margin for line speeds of up to 65mph (105km/h) (Clark et al 2000, p4; WSL 1995, pp3/1-9). The Alcatel Selcab psudo-moving block system has a safety margin of 25m for block lengths of 6.25m (Uebel 1998, pp6-7). The Alstom SACEM system utilises a safety margin of 25m at all speeds for metro services (see Appendix A Interview with Michel Carnot). The author has only found one reference to proposed moving block safety margins on main line railways (for the Alcatel system). This referred to a hypothetical safety margin of 25m for a passenger train at speeds up to 400km/h (249mph) and of 100m for a 2.1 tonne goods train travelling at up to 200km/h (124mph).(Uebel 1998, pp6-8; Appendix A interview with H Uebel).
	Radio range finding	49m	The GE AATC system has a safety margin that varies with location certainty. The defined location accuracy is +-15 feet (4.6m) for each train. In the worst case, this gives a 30m estimated separation error. In practice, the maximum error expected during normal operation is +-10feet (3m), giving an estimated separation error of 6.1m. This is multiplied by a factor of 8 to determine the safety margin required on any movement authority update. Hence, the largest safety margin required would be 73m, whilst the maximum expected in normal operation is 49m (Appendix A Interview with Meryn Parvard).
ATP system trans. delay (U_t).	Movement authority determined trackside	5s	See table 9-1, time includes 3s processing time and 2s for transmission
	Movement authority determined by direct vehicle to vehicle communication	3s	Processing time only required (TCS JPT 2002, System work sheet; Lewis et al. 2002, p27; Appendix A interviews with Sam Macano, Jim Hoelscher and Mark Glover)

Table 9-2: Moving Block System Delay Factors

9.8 OPERATIONAL FACTORS

The operational factors associated with conventional UK colour light signalling were discussed in section 8.7. These must be re-considered to cover the implementation of an ATP system.

In the case of ATP overlay systems, drivers are still expected to comply with the line side signalling. This means that the conventional operational factors still apply on approach to a red aspect. The operational factors on approach to a station stop where the starter is displaying a proceed aspect will also remain the same, since this activity is not supervised by either the underlying signalling or the superimposed ATP system. Therefore, the only factor that may change with implementation of an ATP overlay system is the approach to a speed restriction – which is not controlled by the underlying signalling, but is by the superimposed ATP.

Whilst it could be argued that the speed profile and location information provided in the driver's cab by some ATP systems may enable more accurate control of braking on approach to a speed restriction, it is the author's opinion that this is unlikely to make a significant enough difference to amend the current driving practice of braking to a speed restriction at 90% of the train's nominal full service braking rate, particularly since that information is conventionally provided to the driver via line side signs anyway. It is, therefore, the author's opinion that the same driver behaviour characteristics should be applied to the control of braking under conventional signalling and with an ATP overlay.

The case of in-cab ATP systems is a little more interesting. By nature, these systems must provide adequate information relating to speed profiles and movement authorities to enable safe control of the train's movements by the driver. It remains unlikely that access to such information would enable drivers to significantly improve their performance when braking to a speed restriction. On approach to a station stop or red aspect, the in-cab ATP could only make a difference to driver performance if indication of progress towards the stop location were to be provided on an in-cab display (which is the case in ERTMS systems, for example). In such a case, the displayed information would aid the driver's route knowledge when he/she is judging the train's current location with respect to a target, making mis-judgement less likely. Therefore, since the caution taken in braking rates applied under current professional driving practices is in part to allow for drivers misjudging their location, some increase in the brake rates used could be possible with in-cab ATP. However, the caution adopted in operation also arises from concerns about poor adhesion – which would still be required with in-cab ATP. In light of this, it seems likely to the author that only slightly higher braking rates could be safely utilised by drivers on approach to station stops and limit of movement authorities with in-cab ATP than could be under conventional signalling or ATP overlay. Unfortunately, the author has been unable to either verify this opinion or determine what proportion of the risk perceived in defining professional driving braking rates is due to concern over judgement of locations. In consequence of this, he has also been unable to definitively quantify how much higher braking rates drivers could safely use if his opinion is correct. This would appear to be an area that would warrant further research in the future.

In the absence of further research into the possibility of higher driver braking rates being acceptable under in-cab ATP operation, the author will revert to his belief that any increase that could be safely permitted by the introduction of in-cab ATP would be small and will assume that the conventional professional driving practice should be maintained. This will provide a conservative estimate of any capacity benefits that in-cab ATP may offer.

9.9 SUMMARY

The introduction of an ATP system results in considerable changes in the calculations for headway time. Within this chapter, the author has derived equations for practical headway with intermittent or continuous ATP overlay and in-cab ATP systems.

The impact of fixed block ATP systems on achievable headway can be found by comparison of the equations derived in this chapter and Appendix F with those derived for conventional signalling in chapter 8. This will be considered further in chapter 10.

As noted in section 9.8, it is possible that the driver's normal rate of braking could safely be increased when an in-cab ATP system is applied. Whilst it is unlikely that any such increase would be large, it would contribute to capacity improvement. The potential for, and safe magnitude of, such an increase would appear to be an area warranting further research in the future.

10 MAIN LINE OPERATIONAL IMPACT

10.1 INTRODUCTION

In order to assess the operational impact of different ATP systems, the author developed a series of Visual Basic models in Excel. The assumptions made during this development and the results obtained from the models are discussed within this chapter.

The framework for the models was based on the main line headway equations derived in chapters 8, 9 and Appendix F. Conventional 4-aspect lineside signalling (with no ATP) has been used as a base case for the impact assessment of the main ATP types of intermittent and continuous overlays, fixed and moving block in-cab signalling and relative braking.

The models have been written in visual basic to perform calculations in 1m step intervals, based on data entered in an excel spreadsheet (including speed restriction profiles, signal and track section locations and train performance characteristics). The models determine target speeds based on the data profiles and speed / distance profiles for trains attempting to follow these target speeds. The equations outlined in chapters 8, 9 and Appendix F are then evaluated in order to determine the headway time for each ATP system type as the train passes through the defined section of line.

In the case of conventional lineside signalling and overlay ATP systems, the calculations are performed around each signal location. Fixed block in-cab ATP calculations are performed for each track section change, whilst moving block and relative braking calculations are performed on a metre by metre basis through the defined section of line.

The calculations within the models have been based on the assumptions outlined in Table 10-1. Data preparation for signal separations has been performed in accordance with the UK main line standard for signal separations, GK/RT 0024 Issue 4, assuming mixed traffic operation with 4 aspect signalling on a level gradient (Fleming 2000, p10).

Train length: 200m	Signal sighting time: 8s	Readable distance of signal: 800m
Service Brake Build Up: 3.5s	Full service brake rate: $0.88m/s^2$	Operational brake rate to PSR: $0.792m/s^2$ (90% of full service brake rate)
Operational brake rate to station stop: $0.4905m/s^2$ (5%g)	Acceleration rate: $0.3m/s^2$	Location update interval (Tlu+Tlut): 4s
Train location update error (Tlue): 20m	Signal Replacement Joints 10m after signals	Trackcircuit reset time (Tdc): 3.5s
Signal reset delay on track clear (Tc): 5.5s	Route reset delay (Tr): 8s	
Signal Overlap: 180m	Moving Block Safety Margin: 100m	Intervention margins (Tw+Tp+Tip): 13s
Balise change delay (after Tc): 2.4s (including transmission time)	Continuous ATP Transmission Delay (after Tc): 5s	ATP processing time onboard train (Pt): 1 s
1st balise: 10m before signal	2nd balise: 8s at line speed before signal	3rd balise: 14s at line speed before signal

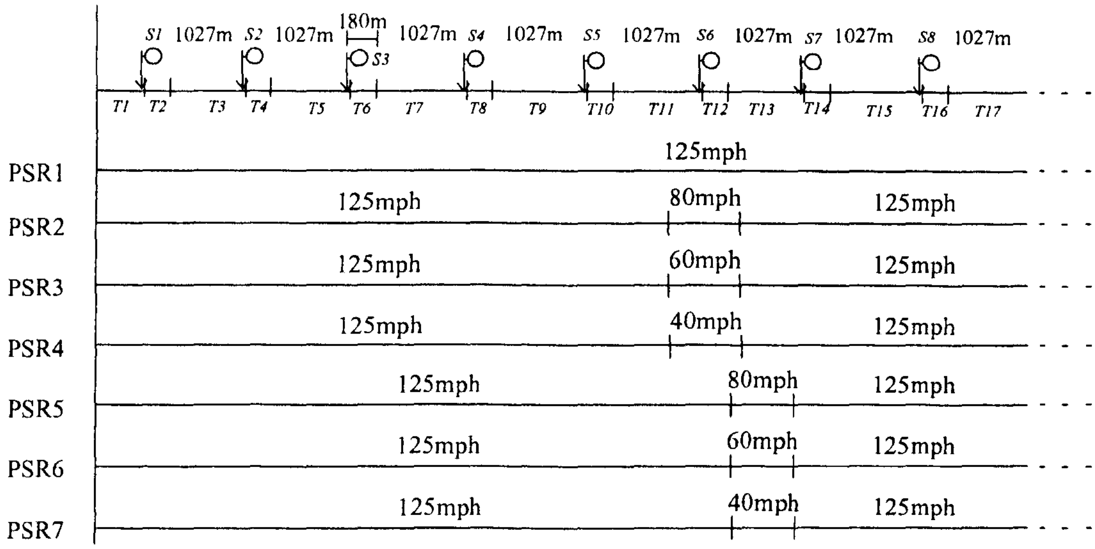
Table 10-1: Assumed Data for Headway Modelling (based on typical figures researched in earlier chapters)

10.2 PLAIN LINE AND SINGLE IN-LINE STATION HEADWAYS

Two main scenarios to consider for speed restrictions were identified in chapter 7. These are a short speed restriction, where signal spacing is not adjusted to the change in permitted speed, and a longer speed restriction for which signal spacing can be adjusted.

10.2.1 SHORT REDUCTION IN PSR

The first of these cases is presented in Figure 10-1. In this case a 500m long restriction (PSR) has been assumed, commencing either 200m before or 265m after a signal, with a station platform stop point located at the same signal location.



PSR Changes: for PSR 2 to 4 the restriction is from 5835m to 6335m. For PSR 5 to 7 it is from 6300m to 6800m

Figure 10-1: Plain Line Signal Layout with Short PSR Reduction (Ideal 125mph Spacing)

10.2.1.1 TECHNICAL HEADWAY

The arrangements shown in Figure 10-1 were implemented in the model, based on Table 10-1, with technical headway braking rates (i.e., use of full service brake rate in all calculations), producing the results shown in Table 10-2. The calculations for continuous in-cab ATP (fixed block) headway assumed use of the same track sections as required for line side signalling.

The results given in Table 10-2 show that with a constant PSR:

- An intermittent ATP overlay system with 1 or 2 balises would produce a 13s (17.2%) increase in headway time when compared to conventional 4-aspect lineside signalling alone. This reduces to 8.4s (11.1%) if a third balise is used;
- The use of continuous radio transmission still increases the headway time, but only by 9.5s (12.6%) for an overlay system or 4.3s (5.7%) with in-cab signalling;
- Moving block operation results in a significant reduction in headway time of 11.4s (15%), whilst the improvement offered by relative braking increases to 43.2s (57%).

The headway obtained with intermittent ATP remains the same whether a one or two balise configuration is assumed, since the ideal ATP "sighting point"[5] is more than 8s running time (the lineside signal sighting time) before the green signal. This is further away from the signal than either balise location, necessitating use of the balise at the previous signal for the ATP sighting point. The third balise location (14s running time before the signal) is slightly before the ideal sighting point and thus imposes a reduced penalty.

---

[5] The ATP "sighting point" is the location at which the trainborne ATP equipment should be updated with less restrictive supervision parameters in order to avoid indication of a need to brake to the driver. This is equivalent to the conventional lineside signalling "sighting point", where a green aspect needs to be clearly visible to the driver to ensure that he/she has time to observe it and react to the clear indication rather than commencing a brake application.

PSR Profile dwell at 6035m	4 Aspect, No ATP	4-Aspect, Intermittent (Balise) Overlay			4-Aspect, Continuous Overlay	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
		1	2	3				
PSR 1	75.4s (47tph)	88.4s (40tph)	88.4s (40tph)	83.8s (42tph)	84.9s (42tph)	79.7s (45tph)	64.0s (56tph)	32.2s (111tph)
PSR 2	95.7s (37tph)	108.8s (33tph)	108.8s (33tph)	104.2 (34tph)	105.3s (34tph)	100.1s (35tph)	82.8s (43tph)	61.5s (58tph)
PSR 3	114.3s (31tph)	128.2s (28tph)	128.2s (28tph)	123.6s (29tph)	124.7s (28tph)	118.6s (30tph)	97.9s (36tph)	80.6s (44tph)
PSR 4	142.8s (25tph)	147.3s (24tph)	147.3s (24tph)	142.8s (25tph)	143.8s (25tph)	138.6s (25tph)	121.0s (29tph)	107.2s (33tph)
PSR 5	94.7s (38tph)	107.7s (33tph)	107.7s (33tph)	103.1s (34tph)	104.2s (34tph)	99.0s (36tph)	82.8s (43tph)	61.5s (58tph)
PSR 6	112.2s (32tph)	125.7s (28tph)	125.7s (28tph)	121.1s (29tph)	122.2s (29tph)	117.0s (30tph)	97.9s (36tph)	80.6s (44tph)
PSR 7	140.6s (25tph)	153.8s (23tph)	153.8 (23tph)	149.2s (24tph)	150.3s (23tph)	145.1s (24tph)	121.0s (29tph)	107.2s (33tph)
30s dwell	219.2s (16tph)	221.8s (16tph)	221.8s (16tph)	219.2s (16tph)	219.2s (16tph)	213.1s (16tph)	195.5s (18tph)	182.8s (19tph)
60s dwell	249.2s (14tph)	251.8s (14tph)	251.8s (14tph)	249.2s (14tph)	249.2s (14tph)	243.1s (14tph)	225.5s (15tph)	212.8s (16tph)

Table 10-2: Limiting Technical Headways, Short PSR Reduction (Ideal 125mph Spacing)

The difference in headway between use of an overlaid ATP or in-cab continuous ATP system is mainly due to the fact that the in-cab system can update a train's movement authority every time that a track section becomes clear, whereas an overlay system is constrained by signal blocks, which typically comprise two track circuit sections. Both systems produce a longer headway than conventional lineside signalling due to the system processing and transmission delays. For example, whilst the lineside signalling calculations assume 8s sighting time, the warning margins for ATP systems have been assumed to total 13s. Similarly, whilst an allowance of 5.5s has been made for a signal aspect change once the required conditions are met, a total of 9.5s has been allowed for detection of the same condition change, transmission and processing of an updated movement authority under continuous ATP.

As changes in PSR are introduced, the variation in trains per hour achieved under different ATP types reduces, until ultimately the introduction of a station stop sees all of the fixed block systems converge to 16tph (for a 30s dwell), while moving block achieves 18tph and relative braking 19tph. It is interesting to note that with severe restrictions, such as PSR 4 or a station stop, the disbenefit caused by introducing fixed block ATP is significantly reduced to between 1 and 3%, rather than 5 to 17%. In the case of continuous in-cab ATP it is actually turned into a 3% improvement. Since any railway will include variations in permitted speed (including station stops), this shows two important facts:

1. The significant reductions in capacity theoretically imposed by fixed block ATP systems during constant speed operation will not actually be experienced on operational railways;
2. The significant increases in capacity theoretically offered by moving block and relative braking during constant speed operation can never be fully utilised.

In the case of PSR3, a significant (and atypical) difference can be observed between the continuous ATP system headways (overlay and in-cab). This occurs because the continuous ATP overlay headway is based around signal locations, the most significant headway being that for which the ATP system targets a stop at 7062m (signal 7). In order to enforce this stop in accordance with the required warning margins, the trainborne ATP system would first indicate the need to reduce speed at 4319m. In contrast to this, in-cab ATP systems are based around track sections. Their most significant headway would be that for which the ATP system targets a stop at 7072m (the end of track 13) and they would first indicate the need to reduce speed at 6194m. This significant difference in indication location arises because a train braking for the speed restriction at 5835m, if it was to continue braking, would come to rest at 7070m (between the overlay and in-cab braking target points). Hence the overlay ATP sighting point is based on a train speed of 125mph, whilst that for the in-cab ATP is based upon 60mph.

It is interesting to note that for fixed block ATP overlay systems, the headway achieved varies with the location of a speed restriction in relation to the signals / track sections (resulting in a difference of up to 1tph for the examples considered in Table 10-2). However, this variation is not particularly significant and will not be considered further.

Perhaps the most significant scenario considered in Table 10-2 is that of a station stop, which produces the most limiting headway for all ATP types. In the case of technical headways, the results given in Table 10-2 show that with a station stop of 30s dwell:

- A 1 or 2 balise intermittent ATP overlay system would produce a 2.6s (1.2%) increase in headway time when compared to conventional 4-aspect lineside signalling alone;
- The use of 3 balises or a continuous radio transmission in an overlay arrangement results in no headway difference, since the ATP headway is actually less than that of the underlying signalling, but drivers are not able to utilise the improvement without disregarding the lineside signals. For an in-cab signalling system this constraint does not apply and a headway time reduction of 6.1s (2.8%) can be obtained;
- Moving block operation results in a significant 23.7s (10.8%) reduction in headway time, whilst the improvement offered by relative braking increases to 36.4s (16.6%).

As the brake rate on approach to a station stop is lower than that on approach to a PSR, the same station stop headway time is achieved for any of the seven PSR profiles.

10.2.1.2 OPERATIONAL HEADWAY – DRIVER PERFORMANCE

Accounting for driver performance and behaviour in the assumed brake rates used to determine headways, the model produces the results shown in Table 10-3. It can be seen that for a constant PSR, where no braking is required, these are the same as the technical headways already discussed. However, reduced braking rates produce different headway results as changes in PSR are introduced.

PSR Profile dwell at 6035m	4 Aspect, No ATP	4-Aspect, Intermittent (Balise) Overlay			4-Aspect, Continuous Overlay	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
		1	2	3				
PSR 1	75.4s (47tph)	88.4s (40tph)	88.4s (40tph)	83.8s (42tph)	84.9s (42tph)	79.7s (45tph)	64.0s (56tph)	32.2s (111tph)
PSR 2	96.0s (37tph)	108.9s (33tph)	108.9s (33tph)	104.7s (34tph)	105.8s (34tph)	100.6s (35tph)	82.8s (43tph)	61.8s (58tph)
PSR 3	114.8s (31tph)	129.1s (27tph)	129.1s (27tph)	124.6s (28tph)	125.6s (28tph)	113.7s (31tph)	91.5s (36tph)	80.8s (44tph)
PSR 4	143.6s (25tph)	148.9s (24tph)	148.9 s (24tph)	144.3s (24tph)	145.4s (24tph)	140.2s (25tph)	119.7s (30tph)	106.8s (33tph)
30s dwell	235.6s (15tph)	235.6s (15tph)	235.6s (15tph)	235.6s (15tph)	235.6s (15tph)	210.2s (17tph)	166.8s (21tph)	161.9s (22tph)
60s dwell	265.6s (13tph)	265.6s (13tph)	265.6s (13tph)	265.6s (13tph)	265.6s (13tph)	240.2s (14tph)	196.8s (18tph)	191.9s (18tph)

Table 10-3: Limiting Operational Headways, Short PSR Reduction (Ideal 125mph Spacing)

The calculations for continuous in-cab ATP (fixed block) headway again assume use of the same track sections required for line side signalling.

For conventional lineside signalling and all of the overlay ATP systems, the reduced braking rates assumed for operational headways increase the time interval required between trains. However, this is not the case for most in-cab ATP, moving block and relative braking results. In several of the results, operating trains at reduced braking performance actually increases the capacity of the line. As these results were not unexpected by the author, they were thoroughly investigated. The results of this investigation vindicated the model and showed that, in some specific instances, headway can be improved by reducing train performance. The explanation draws on two components:

- The ATP system sighting point on approach to a red signal depends on the braking distance, warning margins (between indication of the need to brake and the brake intervention point) and assumed system delays (such as brake build up and processing times). In some instances, the ATP sighting point based on the full service brake rate falls just before the point at which braking must commence to comply with the PSR. Where braking at a reduced rate (90% of the full service brake rate) under operational headway conditions requires brake commencement before this, the ATP sighting point will be based upon the lower PSR speed and thus located significantly closer to the target stopping point (see Figure 10-2 for an example);
- With moving block and relative braking, headway is optimised to the actual speed of operation, rather than a predefined design speed. As already discussed in chapters 2 and 7, the optimal speed for maximising headway with a station stop is in the region of 40 to 50mph, rather than the 125mph operating speed typically used on main line railways. As a result, causing a train to brake at a reduced rate actually reduces the approach speed to the station stop and brings the headway closer to the optimal.

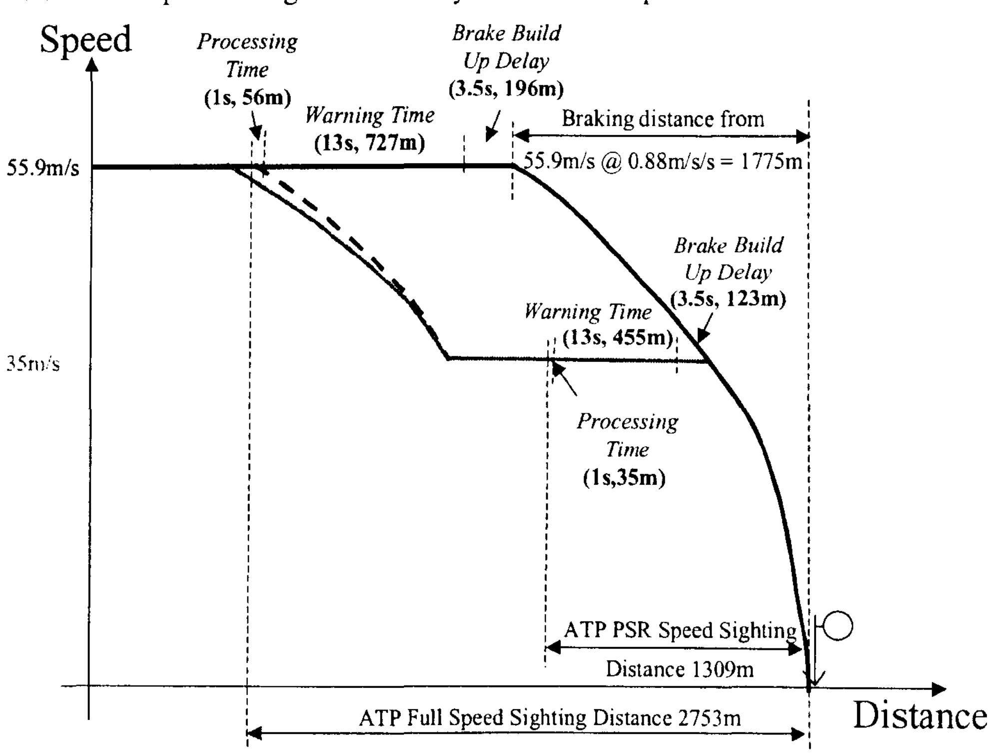


Figure 10-2: Impact of Reduced Driver Braking Rate

Once again, perhaps the most significant scenario considered in Table 10-3 is that of a station stop, which produces the most limiting headway for all ATP types. In the case of operational headways, the results given in Table 10-3 show that for a 30s dwell station stop:

- An intermittent ATP overlay system would produce no change in headway when compared with a conventional lineside signalling system, since the ATP headway is actually less than that of the underlying signalling, but drivers are not able to utilise the improvement without disregarding the lineside signals;
- The use of continuous radio transmission in an overlay arrangement also results in no headway difference. However, with an in-cab signalling system, a headway time reduction of 25.4s (10.8%) can be obtained;
- Moving block operation results in a significant 68.8s (29.2%) reduction in headway time, whilst the improvement offered by relative braking increases to 73.6s (31.2%).

An alternative representation of this scenario can be seen in the speed / distance curve of Figure 10-3. This shows that, for operation along a single line with an in-line station stop, none of the ATP arrangements considered impose a negative impact when the effects of driver performance / behaviour are taken into account. In fact, the potential for capacity improvement with fixed or moving block in-cab signalling becomes highly significant. The improvement to be gained by moving to a relative braking system is, however, far less significant.

In all of the examples so far, continuous in-cab ATP calculations have been based on dividing the track into the sections required for conventional lineside signalling. In practice, this is neither necessary nor optimal for the achieved headway. Simply inserting an additional track section 57m after the station stop location is sufficient to reduce the Continuous In-Cab ATP (fixed block) operational headway for a 30s dwell station stop from 210.2s to 201.2s. Further dividing the overlap track into three (for example, 40 and 80m past the signal) reduces this further to 200.3s. A total of 9.9s (almost 1tph) can be gained by these simple developments. This would have the effect of achieving a headway time reduction of 35.3s (15%) when compared to conventional 4-aspect lineside signalling.

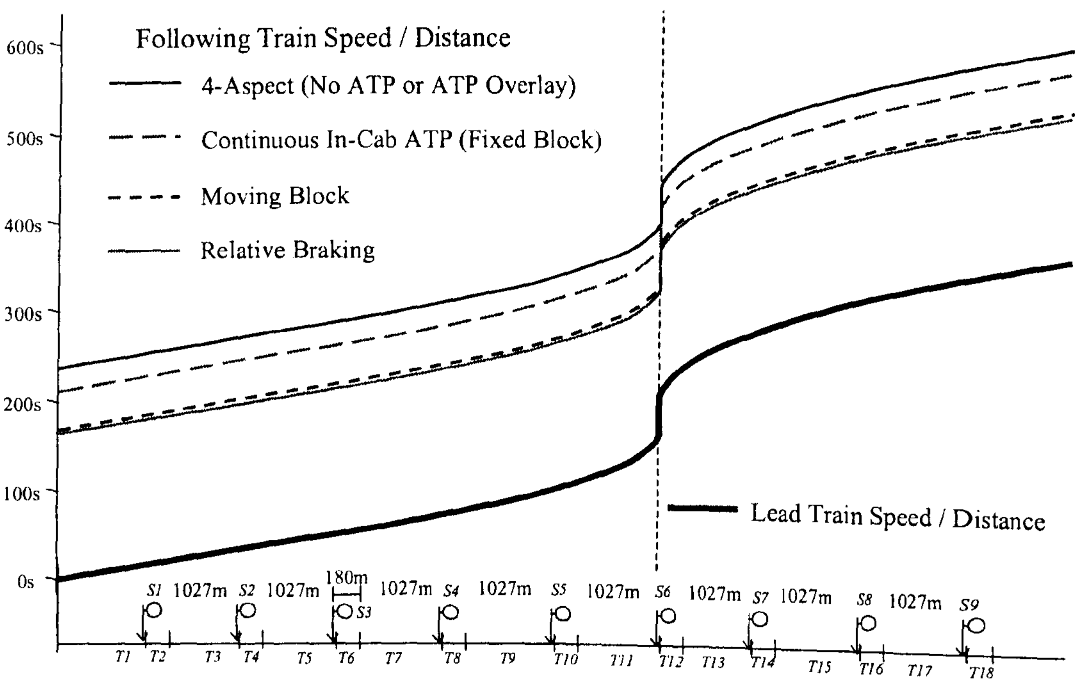


Figure 10-3: Headway Curves for Single In-line Station Stop

10.2.1.3 OPERATIONAL HEADWAY – RECOVERY MARGIN

If operational headway is determined by use of a recovery margin instead of allowing for driver performance, the combined effects of ATP system delays and driver performance could be overcome by allowing a recovery margin equivalent to:

- 17.3% of the underlying 4-aspect technical headway for PSR1;
- 13.8% of the underlying 4-aspect technical headway for PSR2;
- 12.9% of the underlying 4-aspect technical headway for PSR3;
- 4.3% of the underlying 4-aspect technical headway for PSR4 and;
- 7.5% of the underlying 4-aspect technical headway for a 30s dwell station stop.

In all cases this is far less than the 33% margin advocated by the UIC and SRA (Holtzer 1999, p587; Schmid et al. 2002, p7; Steer 2003, p3). In consequence, a railway operating in accordance with the UIC recommendations would not be expected to notice any capacity impact when introducing an ATP overlay system, except in the case of excessive perturbations to the service. This is in keeping with the experience of SNCF and some other European railways, which make large recovery allowances and have experienced no operational impact from the use of intermittent ATP overlay systems. It also offers an explanation of why, in contrast to the European experience, UK main line railways (which use up to 95% of technical capacity) have experienced operational delays following the introduction of intermittent ATP overlay systems (Appendix A interviews with Steve Brown, J Pore and Richard Stanley; Scherp 2003, p2; Wright et al. 2002, p5; discussion following Lundberg 2002, p30).

10.2.1.4 OVERBRAKED SIGNAL SEPARATION

Adjusting the assumed signalling arrangements for 20% overbraking (that is, signal separations 20% longer than required), results in the layout of Figure 10-4. Implementing this arrangement in the model produces the results shown in Table 10-4.

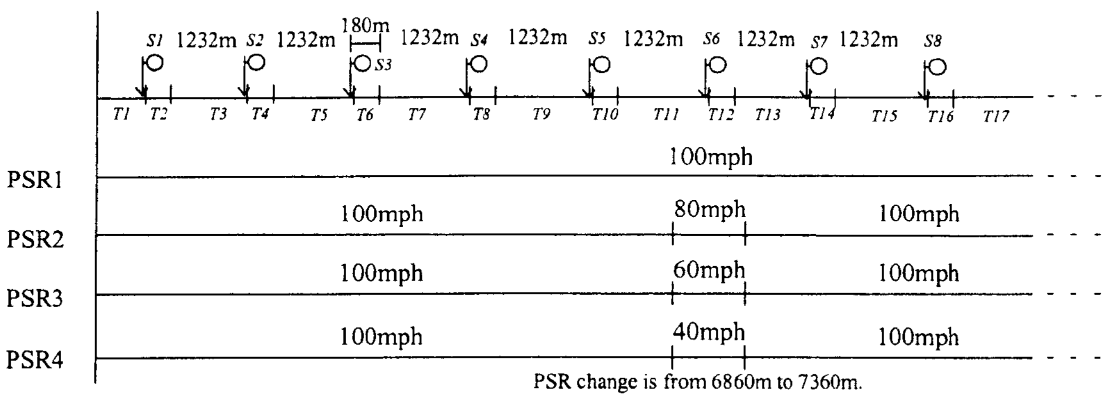


Figure 10-4: Plain Line Signal Layout, Short PSR Reduction (20% Overbraked 100mph Spacing)

Extending the signal separations in this way increases the conventional 4-aspect signalling headway significantly. It also increases all ATP overlay headways when compared to the same arrangement implemented with ideal signal spacing – for both technical and operational headways. The same is true of the fixed block continuous in-cab system results, since they are also based on the track section arrangements used for conventional 4-aspect signalling. However, there is no impact at all on moving block or relative braking systems, significantly improving their performance in relation to the conventional 4-aspect signalling base case.

The results of Table 10-4 include the first difference in headway obtained for 1 and 2 balise intermittent ATP systems. This is because, with the extended signal separation, the ideal sighting point for the ATP system falls between the two balise locations (where it fell between the second and third balise with ideal signal spacing). The 2 and 3 balise systems can then use the second balise for their sighting point, whilst the 1 balise system must use the preceding signals balise. In these circumstances, the benefit gained by the second balise is significant, but the third balise offers no benefit during the planned steady state operating conditions.

PSR Profile dwell at 7060m	4 Aspect, No ATP	4-Aspect, Intermittent (Balise) ATP Overlay			4-Aspect, Continuous ATP Overlay	Continuous In-Cab ATP (fixed block)	Moving Block	Relative Braking
		1	2	3				
Technical Headways								
PSR 1	86.5s (41tph)	103.1s (34tph)	88.9s (40tph)	88.9s (40tph)	88.6s (40tph)	83.4s (43tph)	64.0s (56tph)	32.2s (111tph)
PSR 2	107.4s (33tph)	124.1s (29tph)	109.8s (32tph)	109.8s (32tph)	109.6s (32tph)	104.3s (34tph)	82.8s (43tph)	61.8s (58tph)
PSR 3	127.5s (28tph)	137.9s (26tph)	127.5s (28tph)	127.5s (28tph)	127.5s (28tph)	118.2s (30tph)	97.9s (36tph)	80.8s (44tph)
PSR 4	157.3s (22tph)	157.3s (22tph)	157.3s (22tph)	157.3s (22tph)	157.3s (22tph)	144.9 (24tph)	195.5s (18tph)	182.8s (19tph)
30s dwell	234.1s (15tph)	234.1s (15tph)	234.1s (15tph)	234.1s (15tph)	234.1s (15tph)	219.9s (16tph)	195.5s (18tph)	182.8s (19tph)
Operational Headways								
PSR 1	86.5s (41tph)	103.1s (34tph)	88.9s (40tph)	88.9s (40tph)	88.6s (40tph)	83.4s (43tph)	64.0s (56tph)	32.2s (111tph)
PSR2	107.8s (33tph)	124.5s (28tph)	110.3s (32tph)	110.3s (32tph)	110.0s (32tph)	104.6s (34tph)	82.8s (43tph)	61.8s (58tph)
PSR 3	128.1s (28tph)	138.8s (25tph)	128.1s (28tph)	128.1s (28tph)	128.1s (28tph)	119.1s (30tph)	91.5s (36tph)	80.8s (44tph)
PSR 4	158.3s (22tph)	158.3s (22tph)	158.3s (22tph)	158.3s (22tph)	158.3s (22tph)	146.5 (24tph)	119.7s (30tph)	106.8s (33tph)
30s dwell	252.1s (14tph)	252.1s (14tph)	252.1s (14tph)	252.1s (14tph)	252.1s (14tph)	214.5s (16tph)	166.8s (21tph)	161.9s (22tph)

Table 10-4: Limiting Headways, Short PSR Reduction (20% Overbraked 125mph Spacing)

With the effects of speed restrictions included, the extended signal separations significantly reduce the penalty imposed by overlaying an ATP system (to the extent that the ATP overlay systems offer no penalty under PSR 3 or 4 with a multiple balise system).

Considering further the station stop scenario, in the case of both technical and operational headways the results in Table 10-4 show that with a 30s dwell:

- An intermittent or continuous ATP overlay system would produce no change in headway when compared with a conventional lineside signalling system;
- For technical headway, a 14.2s (6%) reduction can be obtained with in-cab fixed block operation; 38.6s (16.5%) with moving block and; 51.3s (21.9%) with relative braking.
- For operational headway, a 37.6s (14.5%) reduction can be obtained with in-cab fixed block operation; 85.3s (33.8%) with moving block and; 90.2s (35.8%) with relative braking.

Inserting an additional track section 72m after the station stop location is sufficient to reduce the continuous In-Cab ATP (fixed block) operational headway for a 30s dwell station stop from 214.5s to 206.2s. Further dividing the overlap track into three (for example, 42 and 82m past the signal) reduces this slightly further to 204.0s. A total of 10.5s can be gained by these simple developments. This would have the effect of achieving a headway time reduction when compared with conventional 4-aspect lineside signalling of 48.1s (19%).

10.2.2 EXTENDED REDUCTION IN PSR

The second speed restriction case to consider is represented in Figure 10-5. In this case, a 4300m long restriction has been assumed, commencing 1600m before a signal and ending 2300m after it. It is further assumed that the same signal marks a station stop location.

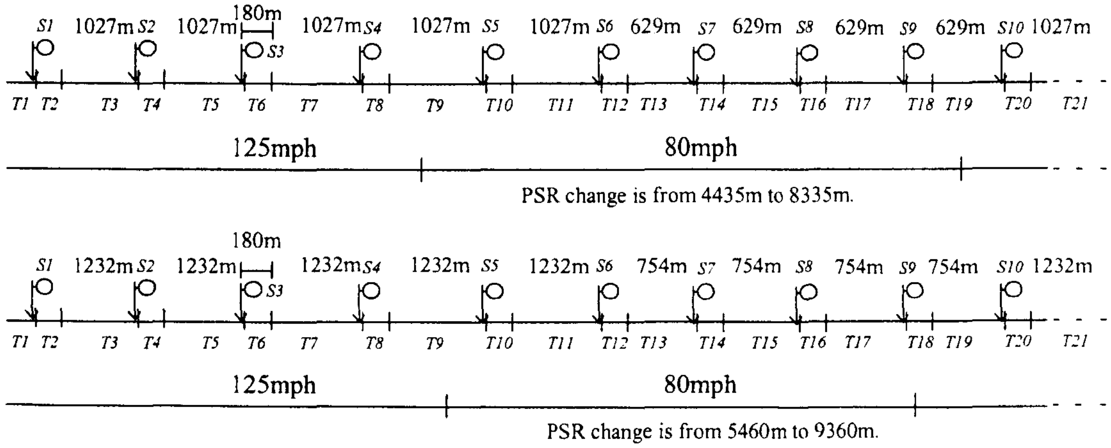


Figure 10-5: Plain Line Signal Layouts with Extended PSR Reduction (Ideal and 20% Overbraked 125 / 80mph Spacing)

Implementing these arrangements in the model produces the results shown in Table 10-5.

Once again, the case of a station stop within the area of consideration has a far more significant impact on headway than the speed restriction itself.

For both technical and operational headways, the long speed restriction at 80mph produces an increase in headway time when compared with the equivalent short speed restriction scenario (PSR2 in Figure 10-1, Table 10-2 and Table 10-3). However, the station stop scenarios show a reduction in headway time with a long PSR. This result can be explained by the fact that the optimal line speed for headway through a single in-line platform is in the region of 40 to 50mph. In the case considered here, the step in speed reduction from 125mph to 80mph occurs before braking to the station stop must commence, enabling reduced track section lengths and improving the headway through the station area, which imposes a more serious constraint for line capacity than the speed restriction itself. This result implies that staged speed restrictions on approach to a station stop would actually produce a better overall capacity for the line (albeit at the expense of a small increase in journey time[6]). The exception to this can be seen in the results for moving block and relative braking under operational braking characteristics. In both of these cases, only a minimal change in headway can be observed between the two sets of operational headway times and, most interestingly, the change is an increase for the long PSR. The explanation for this can be seen in Figure 10-6. With a short PSR, the critical headway conditions under moving block and operational braking rates produce an ATP sighting point at 4014m, for a target stop point at 5920m. This sighting point corresponds to a location already on the operational braking curve to a station stop at 6035m. In the case of an extended PSR, the critical ATP sighting point is at a slightly earlier location (3950m), still on the operational braking curve to a station stop at 6035m and also just before the train begins decelerating for the speed restriction. This location produces a slightly longer headway distance.

---

[6] In the 125 / 80mph restriction example used within this section, the headway reduction for conventional 4-aspect signalling is 17.6s (7.5%) at the expense of a run time increase of 8.8s on approach to the station.

PSR Profile	4 Aspect, No ATP	4-Aspect, Intermittent (Balise) Overlay			4-Aspect, Continuous Overlay	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
		1	2	3				
Technical Headways								
Ideal Spacing	99.5s (36tph)	112.5s (32tph)	112.5s (32tph)	108.4s (33tph)	109.0s (33tph)	103.8s (34tph)	88.5s (40tph)	68.1s (52tph)
Ideal Spacing 30s dwell at 6035m	201.7s (17tph)	215.8s (16tph)	215.8s (16tph)	211.2s (17tph)	212.3s (16tph)	202.9s (17tph)	180.1s (19tph)	172.8 (20tph)
20% Overbraked	110.5s (32tph)	124.7s (28tph)	110.5s (32tph)	110.5s (32tph)	112.7s (31tph)	107.4s (33tph)	88.5s (40tph)	68.1s (52tph)
20% Overbraked 30s dwell at 7060m	215.0s (16tph)	229.3s (15tph)	215.1s (16tph)	215.1s (16tph)	217.2s (16tph)	208.3s (17tph)	180.1s (19tph)	172.8 (20tph)
Operational Headways								
Ideal Spacing	100.0s (36tph)	112.9s (31tph)	112.9s (31tph)	108.4s (33tph)	109.4s (32tph)	104.2s (34tph)	87.8s (41tph)	67.4s (53tph)
Ideal Spacing 30s dwell at 6035m	218.0s (16tph)	219.9s (16tph)	219.9s (16tph)	218.0s (16tph)	218.0s (16tph)	211.2s (17tph)	167.7s (18tph)	162.9 (19tph)
20% Overbraked	111.0s (32tph)	125.2s (28tph)	111.0s (32tph)	111.0s (32tph)	113.1s (31tph)	107.9s (33tph)	87.8s (41tph)	67.4s (53tph)
20% Overbraked 30s dwell at 7060m	231.7s (15tph)	232.2s (15tph)	231.7s (15tph)	231.7s (15tph)	231.7s (15tph)	214.9s (16tph)	167.7s (18tph)	162.9 (19tph)

Table 10-5: Limiting Headways, Extended PSR Reduction (125 / 80mph Spacing)

With a short speed restriction, the higher speeds of operation on approach to the station stop meant that this longer distance produced a shorter headway time than that from 4014m. However, with the longer speed restriction, the train's run time is increased and the headway time produced is marginally longer. The location of the speed restriction in relation to the station stop location is such that the need to brake for the extended speed restriction does not reduce speed at the 3950m sighting point and, therefore, the train does not gain any benefit from the presence of the speed restriction, only the penalty of increased run time.

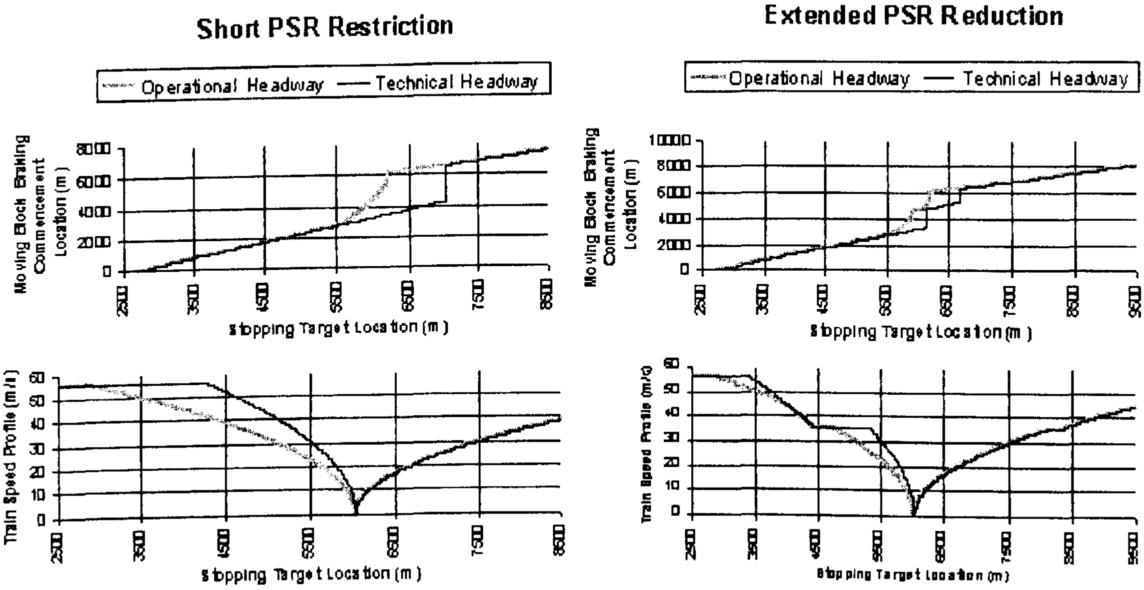
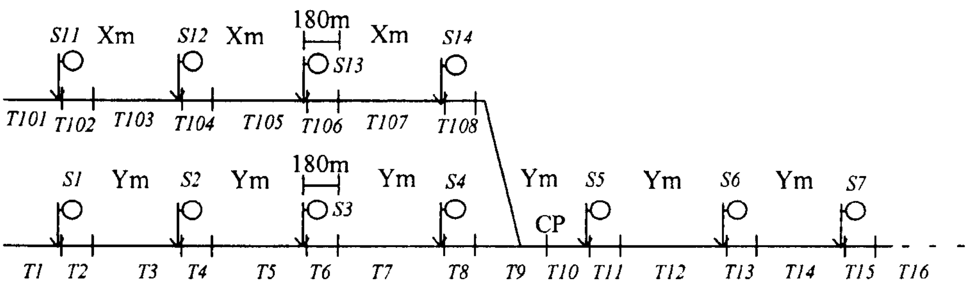


Figure 10-6: Speed Profiles and Braking Commencement / Stopping Target Locations Under Moving Block ATP Algorithms

When analysing examples such as this, it should be noted that it is not just the length of a speed restriction that influences the capacity of a section of line, but also its location with respect to other service characteristics – such a station stop locations.

10.3 CONVERGING JUNCTION HEADWAY

Moving on from plain line sections, the simplest junction arrangement to consider is that of a converging junction. The layout arrangements for ideal and 20% overbraked signal separations represented in Figure 10-7 were entered into the model for converging junctions, producing the results shown in Table 10-6.



Arrangements: Straight Line Speed = 125mph; Converging Line Speed = 100mph
PSR 1: Junction Speed Limit 100mph; T9 = 200m. PSR 2: Junction Speed Limit 80mph ; T9 = 150m
PSR 3: Junction Speed Limit 60mph; T9 = 110m. PSR 4: Junction Speed Limit 40mph; T9 = 80m
Ideal signal spacing: X = 1021m; Y = 1027m; 20% over braked spacing: X = 1225m; Y = 1232m

Figure 10-7: Converging Junction Signal Layout

The dual entries for moving block and relative braking ATP systems in Table 10-6 represent the headway times for 'converging then straight' / 'straight then converging' service patterns. These produce different results due to the junction reset time taking longer than the ATP system update on clearing the junction, which is a constraint on headway for a slower speed converging train following a fast straight train. Where the slow speed train goes first through the junction, the fast train must already be delayed to ensure that it does not catch up before the slow train attains line speed and the junction reset does not form a constraint in this situation.

PSR Profile	4 Aspect	4-Aspect, Intermittent (Balise) Overlay			4-Aspect, Continuous Overlay	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
		1	2	3				
Ideal Spacing, PSR1	88.9s	88.9s	88.9s	88.9s	88.9s	77.8s	62.1s / 69.3s	30.3s / 69.3s
Ideal Spacing, PSR2	99.3s	99.3s	99.3s	99.3s	99.3s	88.3s	72.6s / 79.8s	40.8s / 79.8s
Ideal Spacing, PSR3	117.5s	117.5s	117.5s	117.5s	117.5s	106.5s	90.8s / 98.0s	59.0s / 98.0s
Ideal Spacing, PSR4	144.4s	144.4s	144.4s	144.4s	144.4s	133.4s	117.7s / 124.8s	85.9s / 124.8s
20% Over braked, PSR 1	101.8s	101.8s	101.8s	101.8s	101.8s	81.5s	62.1s / 69.3s	30.3s / 69.3s
20% Over braked, PSR 2	112.2s	112.2s	112.2s	112.2s	112.2s	92.0s	72.6s / 79.8s	40.8s / 79.8s
20% Over braked, PSR 3	130.4s	130.4s	130.4s	130.4s	130.4s	110.2	90.8s / 98.0s	59.0s / 98.0s
20% Over braked, PSR 4	157.3s	157.3s	157.3s	157.3s	157.3s	137.1s	117.7s / 124.8s	85.9s / 124.8s

Table 10-6: Limiting Operational Headways for Converging Junction

It is interesting to note that in all overlay cases the ATP systems would allow trains closer together than the underlying signalling. This is largely due to the small difference in signal spacing permitted for 125mph and 100mph line speeds in accordance with UK main line standards (Fleming 2000, p10). The same speeds of operation have far more significant differences in ATP braking distance. As a result, ATP sighting distances on the converging line are far closer to the supervised movement authority than they are on the straight line, whilst 4-

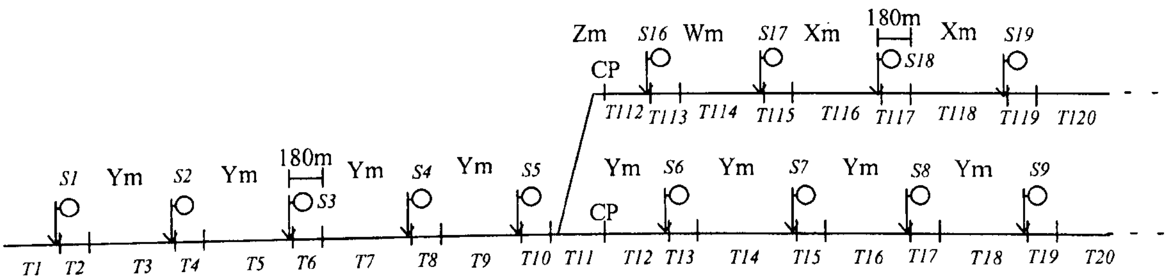
aspect sighting points are in roughly equivalent locations. As a result, the overlay systems do not impact converging junction capacity.

Comparing the converging junction and plain line section headway results of Table 10-2, it can be seen that the junction is more restrictive than a speed restriction of equivalent value under 4-aspect signalling, but less restrictive for all ATP arrangements, with the exception of the relative braking case for a straight train following a converging train (which is significantly constrained by the junction's fixed block effect preventing 'relative' speeds from being utilised). This is despite the arrangement of Figure 10-7 constraining the fixed block in-cab, moving block and relative braking headways to the fixed block junction arrangements, since it has been assumed that there is no separate provision of a point locking track section. In practice, the headway achieved with all three arrangements would be significantly improved by the addition of a track section purely for this purpose, extending from 100m before to 20m after the point tips, for example.

In all cases the converging junction scenario is also far less significant for line capacity than an in-line station stop.

10.4 DIVERGING JUNCTION HEADWAY

A more complex junction arrangement is that of a diverging junction. Of the layout arrangements considered in section 8.3, three (represented in Figure 10-8) were implemented in a model in order to allow for evaluation of ATP impacts. It should be noted that approach release is only applicable with conventional 4-aspect signalling and ATP overlay systems. However, all ATP variants were still analysed in each case due to the difference in speed restriction profiles. As in the previous cases, the model was used to consider both ideal and 20% overbraked signal separations.



Arrangements: Straight Line Speed = 125mph;
1: No Approach Release: Diverging Line Speed = 120mph; Junction Speed Limit 120mph ; T11 = 300m; Z = Y, W = Y
2: Approach Release from Yellow (based on train reaching readable distance, 800m before signal): Diverging Line Speed = 115mph; Junction Speed Limit 115mph; T10 to point tips = 360m; T11 = 610m ; Z = Y, W = X
3: Approach Release from Red(based on train reaching readable distance, 800m before signal): Diverging Line Speed = 100mph; Junction Speed Limit 60mph; T11 = 110m ; Z and W = X
In all cases Ideal signal spacing: X = 1021m; Y = 1027m. 20% over braked spacing: X = 1225m; Y = 1232m

Figure 10-8: Diverging Junction Signal Layout

10.4.1 NO APPROACH RELEASE

The results of implementing the model for the arrangements represented in Figure 10-8 with no approach release are given in Table 10-7.

As implemented in this scenario, the straight / straight case is the same as plain line headway. This is also true, albeit with a modified speed profile, for the diverging / diverging case. Hence the two sets of results are very similar, with the small difference being due to the slight reduction in speed through the junction area.

Scenario	4 Aspect	4-Aspect, Intermittent (Balise) Overlay			4-Aspect, Continuous Overlay	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
		1	2	3				
Ideal Spacing								
Diverging / Straight	75.7s	88.6s	88.6s	84.1s	85.1s	78.2s	76.9s	76.9s
Diverging / Diverging	77.9s	91.5s	91.5s	86.2s	87.3s	82.1s	66.2s	35.7s
Straight / Straight	75.5s	88.4s	88.4s	83.8s	84.9s	79.7s	64.0s	32.2s
Straight / Diverging	75.5s	88.4s	88.4s	83.8s	84.9s	79.7s	76.5s	76.5s
20% overbraked								
Diverging / Straight	86.7s	103.3s	89.1s	89.1s	88.8s	78.2s	76.9s	76.9s
Diverging / Diverging	89.3s	106.7s	91.7s	91.7s	91.4s	86.2s	66.2s	35.7s
Straight / Straight	86.5s	103.1s	88.9s	88.9s	88.6s	83.4s	64.0s	32.2s
Straight / Diverging	86.5s	103.1s	88.9s	88.9s	88.6s	83.4s	76.5s	76.5s

Table 10-7: Limiting Operational Headways for Diverging Junction with No Approach Release

In the straight / diverging case, the actual junction headway times for fixed block systems are lower than the plain line headways approaching it. The values quoted are, therefore, the same as those for the straight / straight case. For moving block and relative braking systems this is not the case, since their headways become constrained by the junction release – imposing a fixed block constraint that increases the headway above that of a plain line section.

This effect can also be seen in the diverging / straight case, where the moving block and relative braking headways are again longer than their diverging / diverging equivalents. In this case, due to the reduced speed profile over the diverging route, the fixed block diverging / diverging headway times through the junction itself are slightly longer than those on approach to it. This makes the fixed block diverging / diverging headway times slightly longer than diverging / straight times.

With ideally spaced 4-aspect signalling, the fixed junction constraints combine with long system reaction and warning delays to make the in-cab systems actually suffer longer headway times when compared to ideally spaced 4-aspect signalling for diverging / straight and, in the case of the fixed block arrangement, also for diverging / diverging. However, with overbraked 4-aspect signalling all overlay ATP systems suffer increased headway times, whilst the three in-cab arrangements are unaffected. As a result, 2.5% overbraking is actually sufficient to make moving block and relative braking headways equivalent to a conventional 4-aspect.

Increasing the level of overbraking to 20% results in:

- Fixed block in-cab ATP producing a 10% improvement for diverging / straight and 3.5% improvement for diverging / diverging train combinations;
- Moving block producing an 11.3% improvement for diverging / straight and 25.8% improvement for diverging / diverging train combinations;
- Relative Braking producing an 11.3% improvement for diverging / straight and 60% improvement for diverging / diverging train combinations.

This highlights a significant lesson: Fixed block arrangements can only be optimised for one train performance characteristic that must be designed into the system from the outset, whilst moving block and relative braking systems can adjust to cope with the performance characteristics of actual trains, even where this changes from one train to another.

Overall, the results of Table 10-7 demonstrate the fact that diverging junctions with no approach release result in a marginally longer headway time than a plain line section under conventional 4-aspect signalling arrangements or any of the overlay ATP systems. With a continuous fixed block in-cab ATP system, the diverging junction offers marginally lower headway times for diverging / straight and marginally higher headway times for diverging / diverging train combinations. With moving block and relative braking, the impact of a diverging / diverging train movement is again a marginal increase in headway time, but the increase resulting from a diverging / straight movement is far more significant. This suggests that the capacity improvements predicted for moving block and relative braking operation on a section of plain line could not be fully utilised in practice.

If the results are compared with those for a converging junction, it is interesting to note that a diverging junction with no approach release actually results in:

i. A smaller capacity impact, due mainly to the minimal speed restrictions imposed on this diverging scenario with 4-aspect signalling and ATP overlay arrangements;
ii. A significantly higher impact with moving block and relative braking for some route combinations, due mainly to the fixed block that the junction represents on approach.

However, this scenario would not appear to be critical to a railway's performance overall, as the impact caused is significantly less than that obtained form large speed restrictions and station stops.

10.4.2 APPROACH RELEASE FROM YELLOW

The results of implementing the model for the arrangements represented in Figure 10-8 with approach release from yellow are given in Table 10-8. In this case, the model assumes that trains brake to stop at a signal using the operational brake rate (5%g), just as in the case of a station stop. The signal is assumed to clear when the train reaches the readable distance (800m before the signal). In order to achieve this, a track circuit commencing a distance equivalent to Tc before the readable distance would be required.

With the introduction of approach release to the underlying signalling, the base case headway is significantly increased. As a result, all of the ATP overlay system headway times also significantly increase. However, with provision of three balises, such that a balise is almost ideally located for the underlying signalling release point, the headway impact of intermittent ATP can be brought to as little as 1.7%, or 1% with 20% over braked signal spacing, through a junction that is approach released from yellow.

Since the in-cab ATP arrangements are limited only by junction speeds, with no signalling imposed speed reductions on approach to the junction protection signal, they all produce lower headway times than an equivalent ideally spaced 4-aspect signalling system for both diverging / straight and diverging / diverging movements. As a result:

- Fixed block in-cab ATP produces a 4.7% improvement for diverging / straight and 7.3% improvement for diverging / diverging train combinations;

- Moving block produces a 6.1% improvement for diverging / straight and 24.6% improvement for diverging / diverging train combinations;
- Relative Braking produces a 6.1% improvement for diverging / straight and 56.8% improvement for diverging / diverging train combinations.

Scenario	4 Aspect	4-Aspect, Intermittent Overlay 1	4-Aspect, Intermittent Overlay 2	4-Aspect, Intermittent Overlay 3	4-Aspect, Continuous Overlay	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
Ideal Spacing								
Diverging / Straight	83.2s	96.2s	96.2s	91.6s	92.7s	79.3s	78.1s	78.1s
Diverging / Diverging	90.9s	107.4s	97.3s	92.2s	94.8s	84.3s	68.5s	39.3s
20% overbraked								
Diverging / Straight	92.1s	108.7s	94.5s	94.5s	94.2s	79.3s	78.1s	78.1s
Diverging / Diverging	101.6s	114.3s	106.4s	102.6s	104.5s	89.0s	68.5s	39.3s

Table 10-8: Limiting Operational Headways for Diverging Junction with Approach Release from Yellow

Increasing the level of overbraking to 20% results in:

- Fixed block in-cab ATP producing a 13.9% improvement for diverging / straight and 12.4% improvement for diverging / diverging train combinations;
- Moving block producing a 15.2% improvement for diverging / straight and 32.6% improvement for diverging / diverging train combinations;
- Relative Braking producing a 15.2% improvement for diverging / straight and 61.1% improvement for diverging / diverging train combinations.

This makes an approach released from yellow junction more significant to overall line capacity than either plain line headway or a junction with no approach release. It also shows that the benefit offered by in-cab systems is higher for this type of junction than for the lower differential speed arrangements represented in the case of no approach release.

10.4.3 APPROACH RELEASE FROM RED

In the case of a junction for which conventional signalling would require approach release from red, the model again assumes that trains brake to stop at a signal using the operational brake rate (5%g) and that the signal would clear when the train reaches the readable distance 800m before the signal. In order to achieve this, a track circuit commencing a distance equivalent to Tc before the readable distance would be required.

The results of implementing the model for the arrangements represented in Figure 10-8 with approach release from red are given in Table 10-9.

With the introduction of approach release from red to the underlying signalling, the base case headway is again significantly increased (as are the headways of all overlay ATP systems). As a result of the significant speed restriction imposed by the junction, this is also the case for in-cab ATP arrangements. However, with the exception of relative braking headway for a diverging / diverging combination, the in-cab system headways (which have no signalling imposed restrictions) do not increase as significantly as the conventional signalling or overlay system headways. It should be noted that:

- The single balise diverging /diverging headway produces such a high headway time because it virtually brings trains to a stand before the signal;
- With provision of three balises, such that a balise is almost ideally located for the underlying signalling release point, the headway impact of intermittent ATP can be brought to as little as 0.5% through a junction that is approach released from red.

Scenario	4 Aspect	4-Aspect, Intermittent (Balise) Overlay			4-Aspect, Continuous Overlay	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
		1	2	3				
Ideal Spacing								
Diverging / Straight	112.4s	125.4s	125.4s	120.8s	121.9s	85.7s	84.5s	84.5s
Diverging / Diverging	128.6s	200.7s	136.2s	129.2s	130.5s	112.7s	93.2s	75.0s
20% overbraked								
Diverging / Straight	123.4s	140.1s	125.8s	125.8s	125.6s	85.7s	84.5s	84.5s
Diverging / Diverging	142.3s	214.4s	149.9s	142.9s	144.2s	118.2s	93.2s	75.0s

Table 10-9: Limiting Operational Headways for Diverging Junction with Approach Release from Red

Overall, with ideally spaced signals:

- Fixed block in-cab ATP produces a 23.8% improvement for diverging / straight and 12.4% improvement for diverging / diverging train combinations;
- Moving block produces a 24.8% improvement for diverging / straight and 27.5% improvement for diverging / diverging train combinations;
- Relative Braking produces a 24.8% improvement for diverging / straight and 41.7% improvement for diverging / diverging train combinations.

Increasing the level of overbraking in the underlying signalling to 20% results in:

- Fixed block in-cab ATP producing a 30.6% improvement for diverging / straight and 16.9% improvement for diverging / diverging train combinations;
- Moving block producing a 31.5% improvement for diverging / straight and 34.5% improvement for diverging / diverging train combinations;
- Relative Braking producing a 31.5% improvement for diverging / straight and 47.3% improvement for diverging / diverging train combinations.

A junction that is approach released from red thus impacts overall line capacity far more than either plain line headway or any other junction arrangement considered. However, it is still less restrictive than an in-line station stop.

It is interesting to note that, under this scenario, the percentage improvements offered by the in-cab ATP systems when compared with conventional 4-aspect signalling are actually higher than those that could be achieved by operation at constant speed through a plain line section. This suggests that where large speed differentials occur at junctions through a route, there is a significant potential for capacity increase through the use of in-cab ATP systems. In such a case, the most balanced headway provision between diverging / straight and diverging / diverging trains is found in the moving block arrangement, suggesting that such an arrangement would be best for ensuring an even traffic density throughout a line.

10.5 MULTIPLE PLATFORM STATION HEADWAY

If a second, off-line, platform is added to a station area, the scenario becomes as shown in Figure 10-9. In this scenario, there are four main route options for any train:

1. Non-stop through the in-line platform;
2. Stopping in the in-line platform;
3. Stopping in the off-line platform;
4. Non-stop through the off-line platform (not considered further).

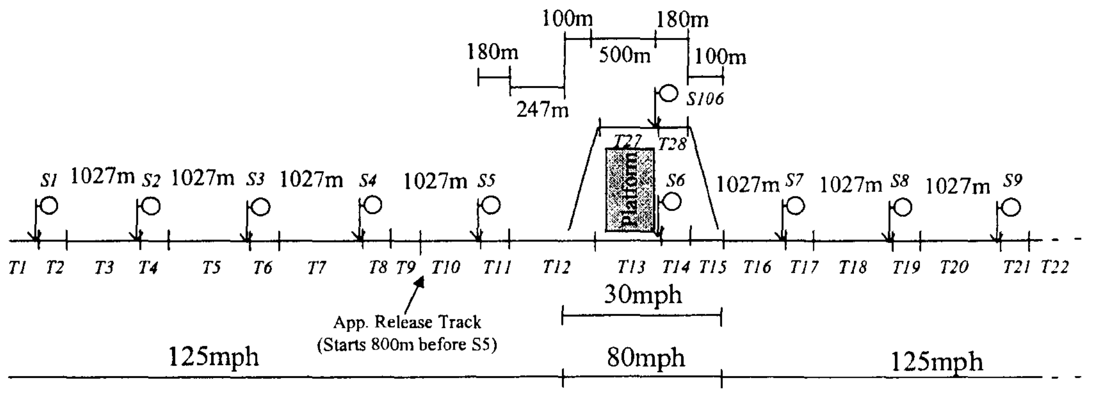


Figure 10-9: Multiple Platform Station Signal Layout, Ideal Spacing

Based on these possibilities, there are two main train-running patterns to consider:

1. All trains stopping;
2. Trains stopping in the off line platform and non-stopping through the in-line platform;

In both cases it is not just the consecutive train headway that is of interest, but also the headway between consecutive trains to the same platform.

The arrangements represented in Figure 10-9 were implemented in a multiple platform station model in order to analyse the impact of differing ATP system approaches.

10.5.1 ALL TRAINS STOP

The results of implementing the model for the arrangements represented in Figure 10-9 with all trains stopping are given in Table 10-10. The main figures are for consecutive trains following the pattern indicated in the left hand column. The supplementary figures (in brackets) indicate the headway required between consecutive trains to the same platform, with a train to the other platform in between.

As would be expected from the previous analysis, all overlay cases for trains following each other into the same platform produce a higher headway time than the underlying signalling. Headway times are highest for a single balise intermittent overlay, reducing with addition of infill or use of continuous overlay. In-cab systems produce headways lower than the conventional 4-aspect signalling case, with the lowest headway achieved for relative braking. This is true both for consecutive trains to the same platform and with a train to the other platform in between them.

The diverging / straight case also produces familiar results. However, the straight / diverging case (which has not been previously analysed) produced some unexpected results – with conventional 4-aspect signalling producing the lowest headway time and in-cab results coming out higher than those for overlay systems. Due to the surprising nature of these results they were investigated in detail to ascertain why they had been produced. It was found that the explanation lies in the inherent speed restriction applied by conventional 4-aspect signalling with approach release from red. Since the train is always supposed to see a red aspect until the release point, the effective sighting point for unrestricted travel is significantly later with approach release than without. Hence the headway for the in-cab system actually ends up slightly longer for the straight / diverging combination where the headway distance is also truncated by route reset, cancelling out the usual headway benefit gained by the reduced look ahead distances required for ATP system headway. This effect is not seen in the diverging / diverging case because the headway distance is not affected by route reset. The in-cab systems therefore still have a shorter overall headway distance than conventional 4-aspect signalling, despite the earlier sighting point.

Scenario	4 Aspect	4-Aspect, Intermittent (Balise) Overlay			4-Aspect, Continuous Overlay	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
		1	2	3				
Ideal Spacing								
Straight / Straight	235.6s (259.6s)	235.6s (298.0s)	235.6s (275.8s)	235.6s (270.7s)	235.6s (265.5s)	204.0s (224.8s)	166.8s (222.3s)	162.0s (222.3s)
Straight / Diverging	99.8s	105.6s	105.6s	101.0s	100.1s	106.0s	104.7s	104.7s
Diverging / Diverging	225.7s (259.6s)	225.7s (298.0s)	225.7s (275.8s)	225.7s (270.7s)	225.7s (265.5s)	218.0s (224.8s)	180.8s (222.3s)	175.9s (222.3s)
Diverging / Straight	159.8s	192.4s	170.2s	169.7s	165.4s	118.9s	117.6s	117.6s
20% overbraked								
Straight / Straight	252.1s (290.4s)	252.1s (335.0s)	252.1s (292.8s)	252.1s (290.4s)	252.1s (290.4s)	214.5s (232.2s)	166.8s (229.7s)	161.9s (229.7s)
Straight / Diverging	106.6s	116.4s	109.0s	106.6s	106.6s	109.6s	108.4	108.4
Diverging / Diverging	241.0s (290.4s)	241.0s (335.0s)	241.0s (292.8s)	241.0s (290.4s)	241.0s (290.4s)	218.0s (232.2s)	180.8s (229.7s)	175.9s (229.7s)
Diverging / Straight	183.8s	218.7s	183.8s	183.8s	183.8s	122.5s	121.3s	121.3s

Table 10-10: Limiting Operational Headways For Two Platform Station With All Trains Stopping In Alternate Platforms

Overall, the results for all trains stopping at a two-platform station show that the in-cab systems do produce lower headways than conventional signalling or overlay systems and that, in the 30s dwell case considered, platform occupancy is no longer a constraint on throughput. For consecutive trains using alternate platforms, the headway required between trains to the same platform improves on that of conventional 4-aspect signalling with ideal signal spacing by:

- 13.4% for fixed block in-cab ATP;
- 14.3% for moving block or relative braking.

The improvement when compared with 20% over braked 4-aspect signalling becomes:

- 20% for fixed block in-cab ATP;
- 21% for moving block or relative braking.

The reader may have noted that the continuous fixed block in-cab ATP headway for straight / straight combinations quoted in Table 10-10 is lower than that quoted in Table 10-3 for the single in-line station stop scenario. The reason for this is the extra track circuits around the junctions, which did not appear in the plain line scenario. In both cases, the headway through the station area is the most significant and the provision of the extra track sections improves the headway obtained for a pure in-cab arrangement. The potential for increasing headway in this way was noted in section 10.2.1.2.

Whilst these results show clear benefits for the implementation of fixed block continuous in-cab ATP, they show only marginal benefits to be gained from introducing moving block – and no additional benefits to be gained from relative braking.

10.5.2 STOPPING OFF LINE, NON-STOPPING IN-LINE

The results of implementing the model for the arrangements represented in Figure 10-9 with trains alternating between stopping off-line and non-stopping in-line are given in Table 10-11. The main figures are once again for consecutive trains following the pattern indicated in the left hand column. The first set of supplementary figures (shown in brackets) indicate the headway required between consecutive trains to the same platform with a train to the other platform in between. The second set of supplementary figures *(shown italic and in brackets)* indicate the headway required between consecutive trains to the same platform with two trains to the other platform in between.

Scenario	**4 Aspect**	**4-Aspect, Intermittent ATP Overlay**			**4-Aspect, Continuous Overlay**	**Continuous In-Cab (fixed block)**	**Moving Block**	**Relative Braking**
		1	**2**	**3**				
Ideal Spacing								
Straight / Straight	90.3s (239.4s)	110.8s (333.4s)	110.8s (262.7s)	106.3s (257.7s)	105.3s (252.4s)	99.4s (205.1s)	84.9s (202.5s)	65.1s (202.5s)
Straight / Diverging	79.6s	92.5s	92.5s	88.0s	87.0s	86.2s	84.9s	84.9s
Diverging / Diverging	225.7s (239.4s) *(329.7s)*	225.7s (333.4s) *(444.2s)*	225.7s (262.7s) *(373.5s)*	225.7s (257.7s) *(364.0s)*	225.7s (252.4s) *(357.7)*	218.0s (205.1s) *(304.5s)*	180.8s (202.5s) *(287.4)*	175.9s (202.5s) *(267.6s)*
Diverging / Straight	159.8s	192.4s	170.2s	169.7s	165.4s	118.9s	117.6s	117.6s
20% overbraked								
Straight / Straight	101.3s (272.8s)	125.5s (324.3s)	111.3s (275.2s)	111.3s (275.2s)	109.0s (272.9s)	103.9s (212.3s)	84.9s (209.9s)	65.1s (209.9s)
Straight / Diverging	89.0s	105.6s	91.4s	91.4s	89.1s	89.8s	88.6s	88.6s
Diverging / Diverging	241.0s (272.8s) *(374.1s)*	241.0s (324.3s) *(449.8s)*	241.0s (275.2s) *(386.5s)*	241.0s (275.2s) *(386.5s)*	241.0s (272.9s) *(381.9s)*	218.0s (212.3s) *(316.2s)*	180.8s (209.9s) *(294.8s)*	175.9s (209.9s) *(275.0s)*
Diverging / Straight	183.8s	218.7s	183.8s	183.8s	183.8s	122.5s	121.3s	121.3s

Table 10-11: Limiting Operational Headways For Two Platform Station With Alternate Stopping off line, Non-stopping In-line

The headways obtained are again consistent with those produced for previous scenarios analysed. Headway times for all routing combinations are highest for a single balise intermittent overlay, reducing with addition of infill or use of continuous overlay. In-cab systems produce headways lower than the conventional 4-aspect signalling case, with the lowest headways achieved for relative braking. This is true both for consecutive trains to the same platform and with one or two trains to the other platform in between them.

The results for stopping (off line) and non-stopping (in line) trains at a two platform station show that, with a 30s dwell, platform occupancy is not a constraint on throughput for conventional signalling, ATP overlay, moving block or relative braking arrangements, but that it is for fixed block-in cab systems. Overall the in-cab ATP systems offer headway improvements for alternate stopping (off line) and non-stopping (in line) trains, when compared with conventional 4-aspect signalling with ideal signal spacing, of:

- 14.3% for fixed block in-cab ATP;
- 15.4% for moving block or relative braking.

The improvement when compared with 20% over braked 4-aspect signalling becomes:

- 22.2% for fixed block in-cab ATP;
- 23.1% for moving block or relative braking.

Once again, these results show clear benefits for the implementation of fixed block continuous in-cab ATP but only marginal benefits to be gained from introducing moving block – and no additional benefits to be gained from relative braking.

For a service pattern with one stopping train (off line) followed by two non-stopping trains (in line), when compared with ideally spaced conventional 4-aspect signalling the in-cab ATP systems offer headway improvements of:

- 7.6% for fixed block in-cab ATP;
- 12.8% for moving block and;
- 18.8% for relative braking.

The improvement when compared with 20% over braked 4-aspect signalling becomes:

- 15.4% for fixed block in-cab ATP;
- 21.2% for moving block and;
- 26.4% for relative braking.

In these cases the benefits to be gained from fixed block in-cab ATP are much smaller (although still significant). Introducing moving block also offers lower benefits in absolute terms, but the differential to be gained by changing from fixed to moving block becomes far more significant, up from around 1% to between 5 and 6%. A similar benefit can also be gained by moving to a relative braking system.

10.6 COMPLEX SCENARIO COMBINATIONS

The analysis presented so far in this chapter has considered only discrete scenarios. However, most real railway operations would in fact consist of a combination of these scenarios. Therefore, it is necessary to consider the overall effect of the various ATP systems when applied to realistic railway layouts. To this end, the author considered three particular complex scenarios to be of interest in determining the overall impact of different ATP types:

1. Traffic from two lines merging at a converging junction. All trains then stopping at a single in-line platform station, before splitting on to two lines at a diverging junction.
2. Repeating the same complex scenario, but replacing the single in-line station with a two-platform station.
3. Repeating the second scenario but with only alternate trains stopping at the multiple platform station.

These scenarios can be built from the discrete scenarios, as shown in Table 10-12.

Route Stage (all stops dwell for 30s)	4 Aspect	4-Aspect, Intermittent (Balise) Overlay		4-Aspect, Continuous Overlay	Continuous In-cab (fixed block)	Moving Block	Relative Braking
		1	2				
Ideal Signal Spacing							
60mph Converging Junction conv / str (str / conv)	117.5s (117.5s)	117.5s (117.5s)	117.5s (117.5s)	117.5s (117.5s)	106.5s (106.5s)	90.8s (98.0s)	59.0s (98.0s)
Plain Line, 125mph PSR	75.4s	88.4s	88.4s	84.9s	79.7s	64.0s	32.2s
Single In-line Station Stop	235.6s	235.6s	235.6s	235.6s	210.2s	166.8s	161.9s
2 Platform Station Stop, all stop. Div/str (str/div)	159.8s (99.8s)	192.4s (105.6s)	170.2s (105.6s)	165.4s (100.1s)	118.9s (106.0s)	117.6s (104.7s)	117.6s (104.7s)
2 Platform Station, stop/non-stop. Div/str (str/div)	159.8s (79.6s)	192.4s (92.5s)	170.2s (92.5s)	165.4s (87.0s)	118.9s (86.2s)	117.6s (84.9s)	117.6s (84.9s)
60mph Diverging Junction div / str (str / div),	112.4s (128.6s)	125.4s (200.7s)	125.4s (136.2s)	121.9s (130.5s)	85.7s (112.7s)	84.5s (93.2s)	84.5s (75.0s)
20% Over Braked Signal Spacing							
60mph Converging Junction conv / str (str / conv)	130.4s (130.4s)	130.4s (130.4s)	130.4s (130.4s)	130.4s (130.4s)	110.2s (110.2s)	90.8 (98.0s)	85.9s (98.0s)
Plain Line, 125mph PSR	86.5s	103.1s	88.9s	88.6s	83.4s	64.0s	32.2s
Single In-line Station Stop	252.1s	252.1s	252.1s	252.1s	214.5s	166.8s	161.9s
2 Platform Station Stop, all stop. Div/str (str/div)	183.8s (106.6s)	218.7s (116.4s)	183.8s (109.0s)	183.8s (106.6s)	122.5s (109.6s)	121.3s (108.4s)	121.3s (108.4s)
2 Platform Station, stop/non-stop. Div/str (str/div)	183.8s (89.0s)	218.7s (105.6s)	183.8s (91.4s)	183.8s (91.4s)	122.5s (89.8s)	121.3s (88.6s)	121.3s (88.6s)
60mph Diverging Junction div / str (str / div)	123.4s (142.3s)	140.1s (214.4s)	125.8s (149.9s)	125.6s (144.2s)	85.7s (118.2s)	84.5s (93.2s)	84.5s (75.0s)

Table 10-12: Limiting Operational Headways for Complex Arrangements

In every case, the single in-line station stop is by far the most restrictive element in terms of capacity. Where one or more single in-line platform station exists on a rail line with all trains scheduled to stop, that station area will determine the overall impact of ATP types on capacity, as shown in Table 10-13. It can be seen that none of the ATP overlay systems actually cause changes to the headways achieved with 4-aspect signalling alone, because the headways applicable to the ATP system alone would actually be lower than the underlying headway. As would be expected from this, all of the in-cab ATP arrangements offer significant headway time reductions. A substantial gain is also made by changing from fixed to moving block – but very little is gained by taking this further to relative braking separation.

ATP Arrangement	Ideal Signal Spacing		20% Over Braked Signal Spacing	
	Headway	% change on 4-Aspect	Headway	% change on 4-Aspect
4 Aspect, No ATP	235.6s	-	252.1s	-
4-Aspect, Intermittent ATP Overlay, 1 Balise	235.6s	No Change	252.1s	No Change
4-Aspect, Intermittent ATP Overlay, 2 Balises	235.6s	No Change	252.1s	No Change
4-Aspect, Continuous ATP Overlay	235.6s	No Change	252.1s	No Change
Continuous In-Cab ATP (fixed block)	210.2s	10.8% less	214.5s	14.9% less
Moving Block	166.8s	29.2% less	166.8s	33.8% less
Relative Braking	161.9s	31.3% less	161.9s	35.8% less

Table 10-13: Overall Impact of ATP Types on Capacity with a Single In-line Station Stop (all trains stop)

Assuming that all stations have two platforms, the most significant scenario element varies with the type of ATP assumed and each train's routing.

A ranking summary of elemental scenario significance, based on the worst-case route combination for each scenario, can be seen in Table 10-14.

<table>
<tr><th rowspan="2">Elemental Scenario Significance forCapacity</th><th rowspan="2">4 Aspect, No ATP</th><th colspan="2">4-Aspect, Intermittent (Balise) ATP Overlay</th><th rowspan="2">4-Aspect, Continuous ATP Overlay</th><th rowspan="2">Continuous In-Cab ATP (fixed block)</th><th rowspan="2">Moving Block</th><th rowspan="2">Relative Braking</th></tr>
<tr><th>1</th><th>2</th></tr>
<tr><td colspan="8">Ideal Signal Spacing</td></tr>
<tr><td>1</td><td colspan="7">Single In-line Station Stop</td></tr>
<tr><td>2</td><td>2 Platform Station</td><td>Diverging Junction</td><td colspan="5">2 Platform Station</td></tr>
<tr><td>3</td><td>Diverging Junction</td><td>2 Platform Station</td><td colspan="3">Diverging Junction</td><td colspan="2">Converging Junction</td></tr>
<tr><td>4</td><td colspan="5">Converging Junction</td><td colspan="2">Diverging Junction</td></tr>
<tr><td>5</td><td colspan="7">Plain Line, 125mph PSR</td></tr>
<tr><td colspan="8">20% Over Braked Signal Spacing</td></tr>
<tr><td>1</td><td colspan="7">Single In-line Station Stop</td></tr>
<tr><td>2</td><td colspan="7">Two Platform Station</td></tr>
<tr><td>3</td><td colspan="5">Diverging Junction</td><td colspan="2">Converging Junction</td></tr>
<tr><td>4</td><td colspan="5">Converging Junction</td><td colspan="2">Diverging Junction</td></tr>
<tr><td>5</td><td colspan="7">Plain Line, 125mph PSR</td></tr>
</table>

Table 10-14: Elemental Scenario Significance by ATP type

In practice, a total of eight routing combinations can be obtained from the base routes:

- Straight / Converging or Converging / Straight at the converging junction;
- Straight / Diverging or Diverging / Straight on entry to the two platform station;
- Straight / Diverging or Diverging / Straight at the diverging junction.

For comparative purposes, the author took all eight routing combinations and determined their most significant element's headway for each ATP type from Table 10-12. The eight values obtained for each ATP type were then averaged to give an indication of the overall capacity impact to be expected. The results of this analysis are given in Table 10-15. It can be seen that in this case the ATP overlay systems all result in an increase on the headways achieved with 4-aspect signalling alone. In the case of an intermittent ATP system with one balise per signal, the increase is highly significant. With two balises or continuous ATP overlay the increase is substantially reduced – but remains at a level that could be significant to a railway operating close to capacity. All of the in-cab ATP arrangements again offer significant headway time reductions. However, it is of interest to note that very little benefit is gained by changing from fixed to moving block, and none at all by taking this further to relative braking separation.

If alternate trains stop in the off-line platform of a two platform station and the trains in between travel non-stop along the straight line (for example, if fast intercity trains are scheduled to overtake slow local trains at the station), the same eight routing combinations can be obtained from the base route / stopping pattern combinations.

ATP Arrangement	Ideal Signal Spacing		20% Over Braked Signal Spacing	
	Average Headway	% change on 4-Aspect	Average Headway	% change on 4-Aspect
4 Aspect, No ATP	141.4s	-	160.1s	-
4-Aspect, Intermittent ATP Overlay, 1 Balise	171.7s	21.4% more	198.0s	18.7% more
4-Aspect, Intermittent ATP Overlay, 2 Balises	150.5s	6.4% more	179.4s	12% more
4-Aspect, Continuous ATP Overlay	148.2s	4.9% more	160.6s	0.3% more
Continuous In-Cab ATP (fixed block)	114.3s	19.2% less	118.4s	26.0% less
Moving Block	111.2s	21.4% less	114.9s	28.2% less
Relative Braking	111.2s	21.4% less	114.9s	28.2% less

Table 10-15: Overall Impact of ATP Types on Capacity When All Stations have Two Platforms per Direction (all trains stop)

For comparison purposes, the author took all eight routing combinations and determined their most significant headway for each ATP type from Table 10-12. The eight values obtained for each ATP type were then averaged to give an indication of the overall capacity impact to be expected. The results of this analysis were found to be exactly the same as those for all trains stopping (given in Table 10-15).

10.7 TARGETED MOVING BLOCK

An introduction to the idea of targeted moving block (that is, applying moving block operation as an overlay on fixed block systems in localised, critical, areas) can be found in Appendix F. There it is concluded that the idea warrants further investigation to determine whether any benefit could be gained by such targeted application.

Analysis of the results already presented in this chapter reveals that it does indeed offer significant potential benefit (see Table 10-16). Localised application to multiple platform stations achieves only a small benefit, of 1.2 to 1.3%, depending on stopping patterns. However, where there are single in-line platforms on the route, applying localised moving block to those areas alone obtains the full capacity benefit to be gained from full moving block application, that isa 20.6% improvement when compared to a fixed block in-cab system. This leaves two questions to be answered:

- Whether the cost and complexity of localised moving block would be significantly different to that of a full moving block system and;
- Whether any potential cost / complexity savings would be worth the loss of full moving block's inherently improved perturbation response.

The answers to these questions fall outside of the scope of this study, but would certainly warrant future investigation.

Moving Block Application area	**Critical Headway (in-line station stop)**	**% change**	**Critical Headway (all stations have 2 platforms per direction)** average values	**% change**
In-cab fixed block	210.2s at station	-	112.5s all stop (102.6s alternate non-stop) at station	-
Station areas only	166.8s at station	20.6%	111.1s all stop (101.3s alternate non-stop) at station	1.2% (1.3%)
Complete application	166.8s at station	20.6%	111.1s stop (101.3s) at station	1.2% (1.3%)

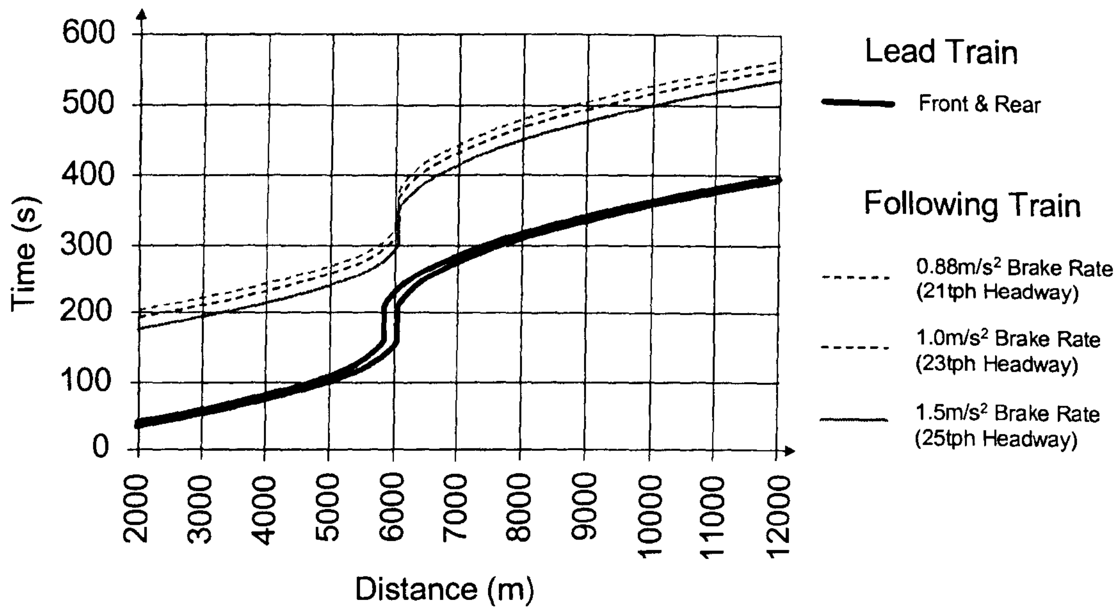
Table 10-16: Impact of Localised Moving Block Application on Capacity

10.8 SENSITIVITY TO BRAKING CHARACTERISTICS

All of the models considered so far have assumed the currently typical UK braking rates of $0.88 m/s^2$ (9%g) for a full service brake application and a $0.49 m/s^2$ (5%g) for a driver's operational brake application. However, the analysis of train braking performance conducted in chapter 5 concluded that, based on practice elsewhere in the world, much higher rates could be reliably applied through use of alternative technology. For the case of high speed main line trains, Table 5-8 shows that service brake rates of up to $1.4 m/s^2$ and emergency brake rates of up to $1.5 m/s^2$ are possible. This raises the question of what impact such braking system enhancements would have on railway capacity.

Clearly such improvements could not be implemented overnight and their potential impact would also be limited if they were brought into effect with existing lineside signals, spaced for the conventional braking rates. Therefore, it would seem likely that the introduction of enhanced braking to the UK would have to coincide with the introduction of moving block signalling and ATP.

Based on this assumption, Figure 10-10 shows the effect that would be gained by varying the ATP intervention rate for a 125mph line speed operation from the current $0.88m/s^2$ up to $1.5m/s^2$, whilst maintaining the drivers brake rate at $0.49m/s^2$, using the single in-line station model analysed in section 10.2. The results suggest that increasing the emergency brake capability would offer significant capacity improvement.



Based on moving block headway with: brake rate to station stop $0.49m/s^2$, 100m safety margin, 200m train length, $0.3m/s^2$ acceleration, 5s transmission delay time, 1s processing time, 4s train location update interval (with 20m accuracy), 3.5s brake build up time and combined ATP warning / intervention margins of 13s.

Figure 10-10: Headway Curves for 125mph Line Speed with Different Emergency Brake Rates

If both the driver's brake rate and ATP braking rates are adjusted for 125mph line speed, the results of Table 10-17 and Figure 10-12 are obtained, revealing interesting, and somewhat unexpected, results:

- That a capacity improvement (of up to 20%) can be obtained with no change in the run-time by increasing the ATP intervention brake rate alone;
- That reductions in run time (of up to 10%) can be obtained by increasing the driver's brake rate alone;
- Increasing the ATP intervention brake rate has no effect on run-time;
- Increasing the driver's brake rate actually reduces the capacity achieved.

ATP Intervention Rate	Driver's Brake Rate	Moving Block Headway	Run-time over 12000m scenario
$0.88m/s^2$	$0.49m/s^2$	166.8s (21tph)	394.9s
$0.88m/s^2$	$0.88m/s^2$	195.5s (18tph)	369.6s
$1.0m/s^2$	$0.49m/s^2$	156.2s (23tph)	394.9s
$1.0m/s^2$	$0.56m/s^2$	166.1s (21tph)	387.8s
$1.0m/s^2$	$1.0m/s^2$	187.9s (19tph)	365.8s
$1.5m/s^2$	$0.49m/s^2$	139.2s (25tph)	394.9 s
$1.5m/s^2$	$0.84m/s^2$	162.1s (22tph)	371.2s
$1.5m/s^2$	$1.4m/s^2$	169.2s (21tph)	357.9s

Table 10-17: Brake Rate Impact on Headway & Run Times for 125mph Line Speed

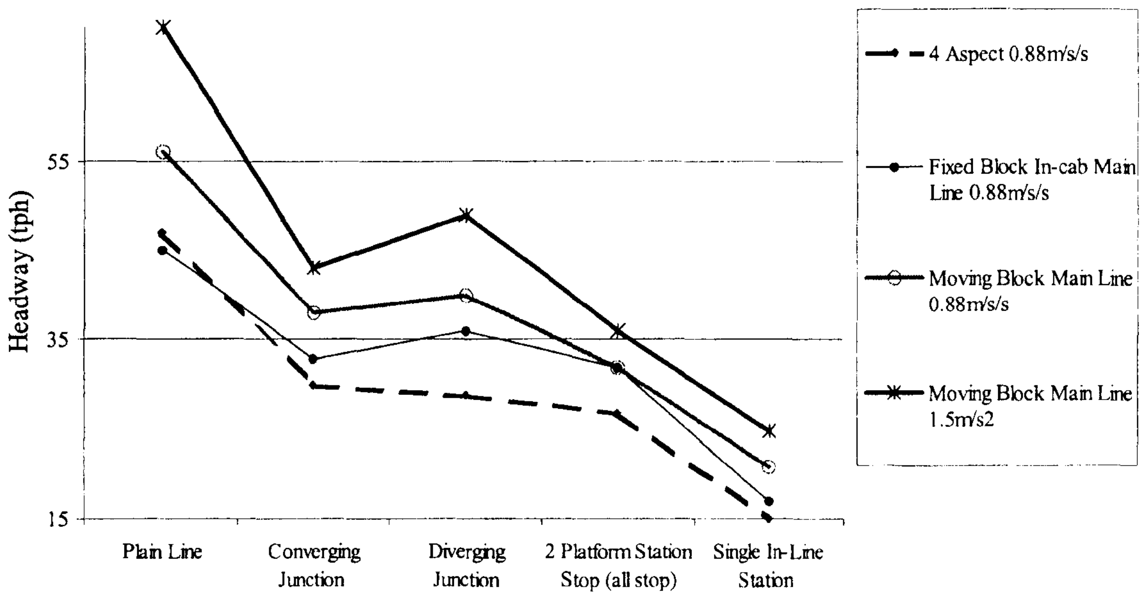


Figure 10-11: Brake Rate impact on Headway for 125mph line speed

In other words, any increase that can be obtained in reliable ATP intervention braking performance offers capacity improvement without detriment to run time but there is a run time / capacity trade off involved in selecting an optimal brake rate for the driver.

When these results are analysed, a clear explanation can be found:

- Increasing the intervention brake rate reduces the safe braking distance from a given speed that must be assumed by the ATP algorithms, whilst increasing the driver's brake rate allows trains to run through the route at higher speeds (braking later to restrictions and station stops). This increases the safe braking distance from any given location on approach to a restriction or station stop, both extending the headway distance and reducing the time required to run through that distance. At high speeds, where the majority of the headway distance is determined by the braking distance (increasing with the square of the speed), this has an overall negative impact on headway time. i.e., the extended braking distance resulting from the higher speeds of operation mean that the headway distance takes longer to traverse, despite the speeds of travel being higher.
- A low driver's brake rate coupled with a high ATP intervention brake rate results in the full speed ATP sighting point falling within the service braking curve. As a consequence, the ATP system will not need to indicate a speed reduction that would in any way interfere with the driver's normal station approach, regardless of the presence of a train ahead. The actual ATP sighting point in this scenario will occur when the driver's service brake curve intersects with the ATP braking curve. This gives a significantly reduced headway distance that will be traversed at lower speeds. At high speeds, where the majority of the headway distance is determined by the braking distance (increasing with the square of the speed), this has an overall positive impact on headway time.

An example can be found in Figure 10-12. The steeper solid line represents a train's braking curve to rest in a station, based on $1.5m/s^2$ deceleration. If a driver were to approach with the same braking rate, the brake would need to be applied for a station stop 1041m before the stop point. At that brake application point, the ATP supervision algorithm (the top dotted line) would look much further ahead than the station. The resulting ATP headway distance would therefore include the full deceleration curve, platform dwell and a substantial part of the acceleration curve after the station. That is, 37s + 30s + 92s = 159s, assuming a 30s dwell, 200m train length, 100m safety margin & $0.3m/s^2$ acceleration rate.

In contrast to this, a driver braking to stop in the station at $0.49m/s^2$ would follow the shallower solid line and commence brake application 3189m before the station stop point. The train would then only actually be travelling 41.7m/s by the time that it reached the full speed ATP brake application point (1041m before the station stop point). The ATP would, therefore, no longer need to initiate a brake application at that point. In fact, it would not need to do so until 560m before the station stop point (following the lower dotted curve). The headway time would then include the rest of the braking distance and the station dwell, but only a small part of the acceleration curve on departure. That is, 65s + 30s + 32s = 127s, assuming a 30s dwell, 200m train length, 100m safety margin & 0.3m/s acceleration rate. i.e. 32s lower in this specific case.

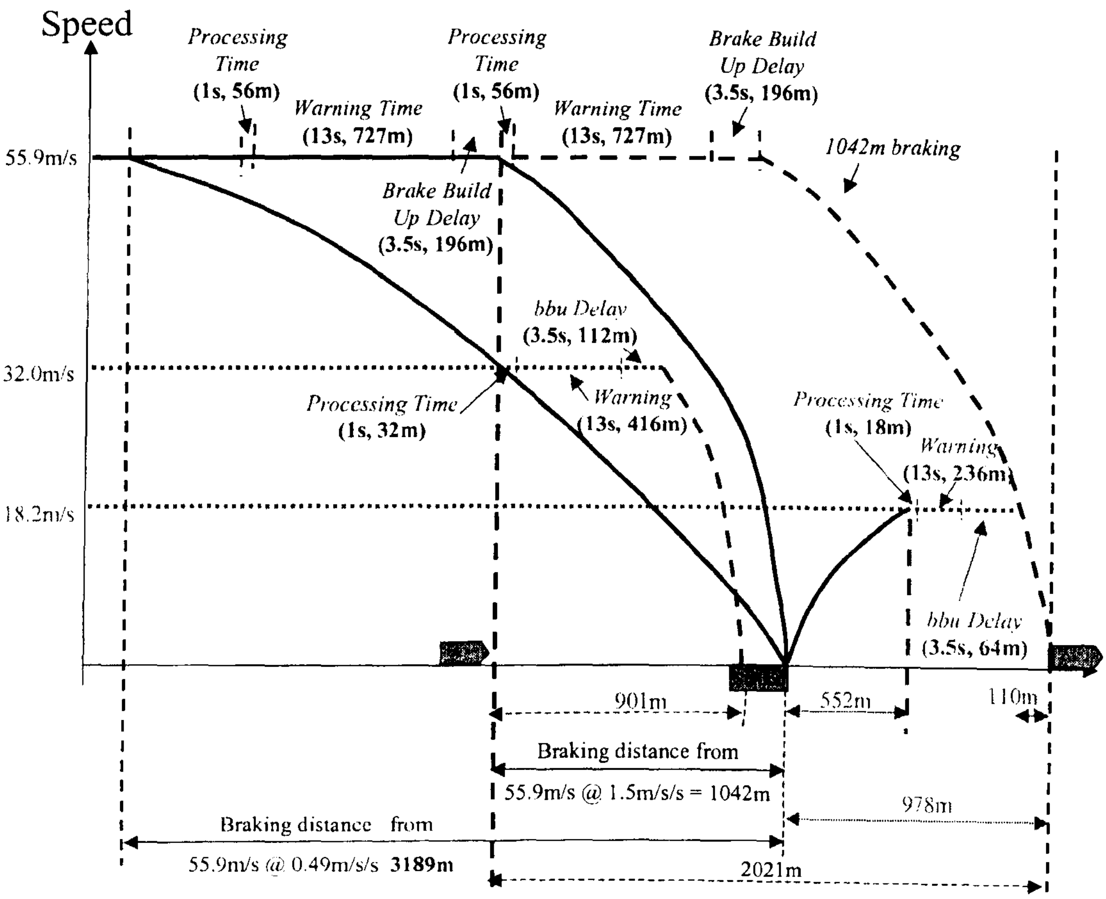


Figure 10-12: Headway Impact of Braking Rates on High Speed Approach to a Station Stop

Enhanced braking performance clearly offers significant potential for future capacity improvement. Such enhancement would be far more significant than that to be gained by developing moving block into relative braking separation.

Whilst not directly a part of the ATP system, it is only to be expected that rollout of trains with enhanced braking capability would be by gradual introduction rather than a big bang approach. This would make optimisation of any fixed block system impossible until all trains had been upgraded. However, with moving block the enhanced performance characteristics of any upgraded train could be utilised as soon as it began operating and some capacity improvement could be expected from commencement of the rollout. This is a potential benefit of moving block ATP that should be considered when determining an optimal strategy for ATP implementation.

10.9 MODEL VALIDATION

At each stage of development the models were tested in order ensure that the implementation was correctly calculating the equations. The testing took two forms:

- The model was programmed to output the location, target speed, actual speed and time variables for each 1m step. The same calculations were performed by hand, calculating run times and braking distances by use of Newton's laws of motion for each assumed station stop or speed restriction. The hand calculated values were then compared with the model outputs to show that the train speed / distance and time / distance profiles produced agreed. This exercise was then repeated for several combinations of speed profile in order to ensure that the agreement was not just a coincidence;
- The overall headway time results were analysed in order to detect trends and identify the impact of different ATP systems. Whenever an unexpected result was noticed, or an explanation for a trend was not already known, the model was re-programmed to output variables such as sighting point, headway start times, etc. and the outputs were investigated to identify explanations. In order to confirm explanations, calculations for specific instances were then repeated by hand.

Several errors in the original model code were identified in this way. In particular:

- The original code for ATP sighting point calculation began at the movement authority and calculated backwards to find the point of brake commencement. Some unexpected results revealed that the ATP look ahead algorithm could actually have several 'brake commencement' points. The furthest out from the movement authority was required, but the model was finding the closest in. The model code was rewritten to overcome this by starting to look for the ATP sighting point from location 0m and moving the search towards the movement authority instead;
- The original code for calculating braking curves was adding a significant delay to station stop braking times by neglecting the fact that the model's step interval was 1m, whilst the braking distance is generally not an exact multiple of that. This was observed when the time / distance outputs were found not to match the hand calculations and the code was re-written to allow for the actual braking distance.

10.10 SUMMARY

The analysis of elemental scenarios within this chapter has revealed the theoretical capacities of each ATP type, enabling comparison between them for different situations. A summary of the findings of this analysis can be seen in Table 10-18 and Figure 10-13. In all cases there is a general trend that can be observed:

i. ATP overlay systems reduce capacity, the impact reducing as the update frequency is increased;

ii. In-cab systems increase capacity, the impact increasing from fixed to moving block and from moving block to relative braking.

Elemental Scenario	4 Aspect	4-Aspect, Intermittent (Balise) Overlay 1	4-Aspect, Intermittent (Balise) Overlay 2	4-Aspect, Continuous Overlay	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
Ideal Signal Spacing							
Plain Line, 125mph PSR	47tph	40tph	40tph	42tph	45tph	56tph	111tph
Single In-Line Station Stop (30s dwell)	15tph	15tph	15tph	15tph	17tph	21tph	22tph
60mph Converging Junction	30tph	30tph	30tph	30tph	33ph	38tph	45ph
60mph Diverging Junction	29tph	22tph	27tph	28tph	36tph	40tph	45tph
Multiple Platform Station (all trains stop)	27tph	24tph	26tph	27tph	32tph	32tph	32tph
20% Over Braked Signal Spacing							
Plain Line, 125mph PSR	41tph	34tph	40tph	40tph	43tph	56tph	111tph
Single In-Line Station Stop	14tph	14tph	14tph	14tph	16tph	21tph	22tph
Converging Junction	35tph	35tph	35tph	35tph	44tph	54tph	72tph
Diverging Junction	27tph	20tph	26tph	26tph	35tph	40tph	45tph
Multiple Platform Station (all trains stop)	24tph	21tph	24tph	24tph	31tph	31tph	31tph

Table 10-18: Operational ATP System Throughput by Elemental Scenarios

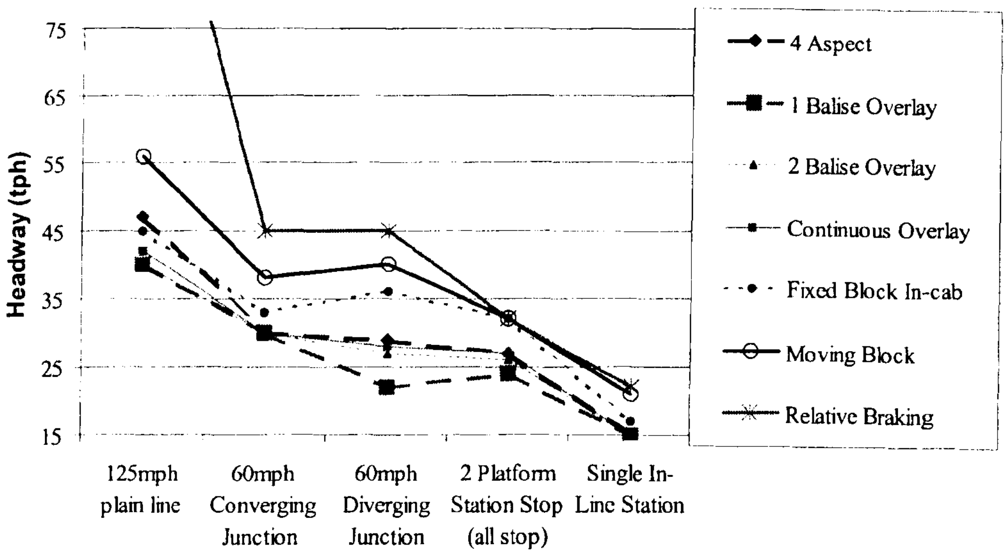


Figure 10-13: Operational ATP System Throughput by Elemental Scenarios (ideal Spacing)

As interesting as the analysis of individual element scenarios may be, it is in the complex scenario combinations that the true impact of different ATP system types on an overall line's performance is to be seen. As already discussed in section 10.6 a single in-line station stop, if present, is always the limiting element scenario for capacity. In such situations, no capacity impact is predicted for ATP overlay systems, an increase in capacity of 11 to 15% is expected from fixed block in-cab systems, while moving block would provide an increase in the region of 29 to 34% and relative braking of 31 to 36%, depending on underlying 4-aspect signal spacing. This potential benefit for steady state capacity improvement from moving block and relative braking could also be obtained by limited application to the station areas only, with the rest of the line fitted for fixed block in-cab operation.

If all stations include two platforms per direction, the limiting scenario varies with the ATP type applied. In this case, the use of ATP overlay systems would be expected to impact capacity by between 0.3 and 21%, depending on 4-aspect signal spacing and the frequency of supervision criteria updates. A fixed block in-cab system would be expected to give an increase in capacity of 19 to 26%, moving block and relative braking of 21 to 28%, again depending on the 4-aspect signal spacing. In such a case, the application of localised moving block would appear to offer insufficient benefit to warrant implementation, whilst relative braking would offer no benefit at all.

Clearly, the findings demonstrate a good case for the implementation of ATP in a continuous in-cab form. On grounds of capacity impact alone, there is also a clear advantage of moving from fixed to moving block (in the region of 20% with in-line station stops and 3% otherwise). In addition to this capacity benefit, it has already been noted in section 4.2.4 that significantly less trackside equipment is required, increasing reliability, reducing costs and reducing risks to staff. With moving block there is also the potential optimising performance under perturbations in accordance with the actual conditions found, rather than the design stage limitations inherent in the use of fixed block sections.

The case for transition from moving block to relative braking is not so clear. Whilst it would offer an increase in the region of 3% with in-line station stops, there would be no benefit if all stations included two platforms per direction. However, as noted in Appendix F4.3.1, a significant level of extra risk would be introduced, specifically due to compound collisions. The technology required is also untried, unlike moving block – for which numerous implementations have been proven in operation on metro railways. In the light of these concerns and the small potential benefit, it is the author's opinion that relative braking does not currently offer a viable, or particularly attractive, option for main line railways.

These findings are broadly in line with the few published predictions of the capacity improvement to be gained by introduction of overlay and fixed block in-cab ERTMS solutions. The UK based ERTMS Programme team has predicted, for example:

- ERTMS Level 1, no infill (a specific case of 1 balise intermittent ATP) 0 to 30% capacity reduction across the network;
- ERTMS Level 1, single infill (a specific case of 2 balise intermittent ATP) 0 to 15% capacity reduction across the network;
- ERTMS Level 2 overlay (a specific case of fixed block continuous overlay) broadly neutral impact of 0 to 5% capacity reduction across the network;
- ERTMS Level 2, in-cab only (a specific case of continuous fixed block in-cab ATP) potential for up to 10% capacity increase.

  (Waboso 2002, p47)

The Dutch BEV 21 has predicted higher capacity increases of 10 to 25% for an ETCS Level 2 in-cab system (Holtzer 1999, p587).

The findings for moving block are also within the lower bounds of the 30 to 50% increase over conventional fixed block-signalling systems suggested in the only numerical prediction that the author has discovered (Wang et.al. 1993, p153).

11 METRO OPERATIONAL IMPACT

11.1 INTRODUCTION

Unlike main line railways, it seems highly unlikely that future metro signalling schemes would consist of a 'conventional' arrangement, whether 2 aspect lineside signalling with trainstop protection or track circuit based speed code systems such as FS2000. For example, all schemes currently being proposed for LUL lines as a part of the Public Private Partnership arrangements will see a form of in-cab signalling and comprehensive ATP – this combination being necessary if both the safety and capacity requirements of a modern metro are to be met. The only question relates to the type of transmission arrangement for the ATP data, with the choice being between:

- A radio based transmission system, with train location determined by trackside equipment (equivalent to ERTMS Level 2) or;
- A radio based transmission system, with trainborne location and integrity proving systems (equivalent to ERTMS Level 3).

As a result of this, the author has only considered three arrangements when analysing metro capacities: continuous in-cab ATP; moving block and relative braking. In order to assess the operational impact to be expected from these arrangements, the Visual Basic models developed for main line scenarios were used, with the data amended to represent typical metro conditions. The results are discussed within this chapter.

The calculations within the models have been based on the assumptions outlined in Table 11-1. It is important to note that metro train services generally consist of a homogeneous rolling stock fleet, with all trains following the same service pattern, stopping at every station. The process of signalling design therefore follows a fundamentally different approach on metros compared to main line railways and signalling equipment is generally located where required to achieve the desired headway, with no requirement to keep signals or track sections distributed evenly along the line.

Train length: 100m	Service Brake Build Up: 3.5s	Acceleration rate: 1.0m/s^2
Moving Block Safety Margin: 25m	Train location update error (Tlue): 20m	Route reset delay (Tr): 8s
Full service brake rate: 1.15m/s^2	Operational brake rate to station stop or PSR: 1.03m/s^2 (90% full service brake rate)	Trackcircuit reset time (Tdc): 3.5s
Location update interval (Tlu+Tlut): 4s	Continuous ATP Transmission Delay (after Tc): 5s	Line Speed: 40mph (60km/h).
Intervention margins (Tw+Tp+Tip): 13s	ATP processing time onboard train (Pt): 1 s	Station dwells 30s
Signal reset delay on track clear (Tc): 5.5s	Signal Replacement Joints 10m after signals	

Table 11-1: Assumed Data for Headway Modelling (based on typical figures researched in earlier chapters)

11.2 PLAIN LINE AND SINGLE IN-LINE STATION HEADWAYS

Plain line headway on a typical metro is limited only by the quantity of signalling equipment that is required to provide an equivalent capacity to that achievable through stations. A typical metro in-line station is represented in the scenario of Figure 11-1. The signalling arrangements shown were implemented in the model, based on both the technical and operational braking rates identified in Table 11-1, producing the results shown in Table 11-2. It has also been assumed in this modelling exercise that plain line route is divided into track sections of 100m for the fixed block in-cab signalling arrangement.

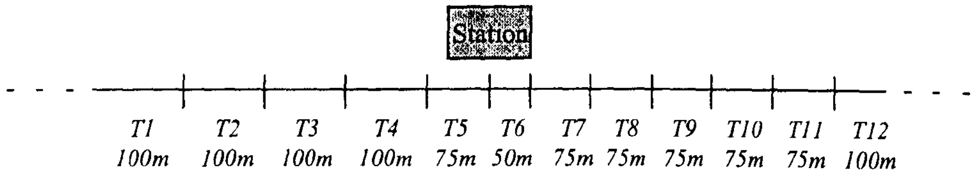


Figure 11-1: Single In-line Station Layout

Brake Rate	**Plain Line (100m track sections)**			**Station Stop (Figure 11-2)**		
	Continuous In-Cab ATP (fixed block)	Moving Block	Relative Braking	Continuous In-Cab ATP (fixed block)	Moving Block	Relative Braking
Full Service Brake ($1.15m/s^2$)				94.5s (38tph)	89.1s (40tph)	84.9s (42tph)
Operational ($1.035m/s^2$)	42.1s (85tph)	30.9s (116tph)	23.1s (155tph)	95.3s (37tph)	90.0s (40tph)	85.8s (41tph)
Reduced / main line equivalent ($0.49m/s^2$)				104.9s (34tph)	99.6s (36tph)	95.4s (37tph)

Table 11-2: Limiting Headways for In-line Station Stop (40mph Line Speed)

The results given in Table 10-2 show that, with a constant PSR, moving block provides a 26.6% reduction in headway time and relative braking a 45.1% reduction, when compared with the fixed block in-cab arrangement.

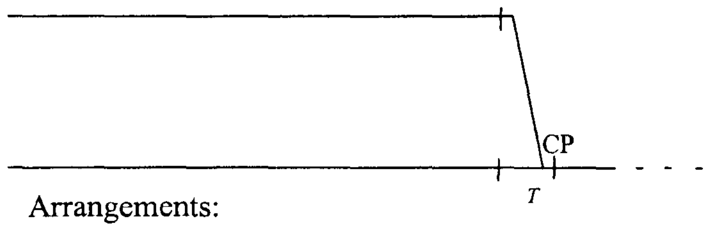
With the introduction of an in-line station stop the headway time is significantly increased, reducing the benefit to be gained by use of moving block or relative braking train separations. With the use of full service braking rates, moving block operation results in a small but significant reduction in headway time of 5.4s (5.7%), whilst the improvement offered by relative braking is 9.6s (10.2%). Under operational braking rates, moving block reduces headway by 5.3s (5.6%) and relative braking by 9.5s (10.0%). Overall, such a small change between full service and operational brake rates increases headway by approximately 1% and produces a negligible difference in the three system's relative performance. If a larger reduction to 5%g is imposed on braking performance, as used under main line professional driving practice, moving block reduces headway by 5.3s (5.1%) and relative braking by 9.5s (9.1%), again an almost negligible difference in the three system's relative performance. However, in this case the change increases each headway value by around 12%. It is interesting to contrast these increases with the headway time reductions obtained by reducing the driver's brake rate at higher main line speeds. Since metro operating speeds are near to the optimal values for headway (see Figure 2-9), the slower station approach speeds imposed by reduced braking rates no longer offer a capacity benefit.

11.3 CONVERGING JUNCTION HEADWAY

The basic converging junction arrangement represented in Figure 11-3 was entered into the respective model, producing the results shown in Table 11-3. As in the plain line scenario already considered, it has also been assumed in this modelling exercise that the plain line route is divided into track sections of 100m for the fixed block in-cab signalling arrangement.

The 'converging then straight' / 'straight then converging' service patterns again produce different results due to the junction reset time taking longer than the ATP system update on clearing the junction. This is a constraint on headway for a slower speed converging train following a fast straight train. Where the slow speed train goes first through the junction, the fast train must already be delayed to ensure that it does not catch up before the slow train attains line speed and the junction reset does not form a constraint.

Comparing the converging junction and plain line section headway results of Table 11-2, it can be seen that under a typical metro arrangement the junction is more restrictive than plain line, but far less restrictive than an in-line station stop. This result is the same as that for the higher speed main line scenario.



Arrangements:
Straight Line Speed = 40mph; Converging Line Speed = 40mph
Junction Speed Limit 30mph; T = 70m
All other track sections for fixed block arrangement = 100m

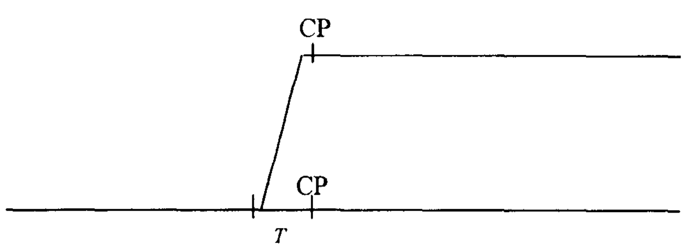
Figure 11-3: Converging Junction Layout

PSR Profile	Continuous In-Cab ATP (fixed block)	Moving Block	Relative Braking
Converging / Straight	44.9s	36.0s	26.2s
Straight / Converging	51.2s	42.4s	32.6s

Table 11-3: Limiting Operational Headways for Converging Junction

11.4 DIVERGING JUNCTION HEADWAY

The basic diverging junction arrangement represented in Figure 11-4 was entered into the appropriate model, producing the results shown in Table 11-4. As in the plain line scenario already considered, this modelling exercise assumed sections of plain line route to be divided into track sections of 100m for the fixed block in-cab signalling arrangement.



Arrangements:
Straight Line Speed = 40mph; Converging Line Speed = 40mph;
Junction Speed Limit 30mph; T = 70m;
All other track sections for fixed block arrangement = 100m

Figure 11-4: Diverging Junction Layout

As implemented in this scenario, the straight / straight case is the same as plain line headway. This is also true, albeit with a modified speed profile, for the diverging / diverging case. The difference in headway time that results from the speed profile change is more significant than that found for main line speeds (see Table 10-7). This is partly explained by considering the fact that the main line scenario for a junction with no approach release assumed a speed differential of 5mph rather than that of 10mph assumed in this case. The other significant difference relates to the line speeds. Under moving block and relative braking, reducing speeds from 125 to 120mph moves the speed profile closer to maximum capacity speeds (see Figure 2-9) and limits the impact of the speed reduction. However, reducing speeds from 40 to 30mph moves them away from the ideal and thus increases the capacity impact of the speed restriction.

Overall, the results of Table 11-4 demonstrate the fact that, at metro speeds, diverging junctions are far more significant for capacity than plain line sections and slightly more significant than converging junctions. The results further show that large capacity gains can be achieved by changing from fixed to moving block and again from moving block to relative braking. However, these gains are not as high as for the plain line scenario, suggesting that once again the capacity improvements predicted for moving block and relative braking operation on a section of plain line could not be fully utilised in practice.

PSR Profile	Continuous In-Cab ATP (fixed block)	Moving Block	Relative Braking
Diverging / Straight	60.6s	56.4s	56.4s
Diverging / Diverging	54.8s	46.7s	41.7s
Straight / Straight	42.1s	30.9s	23.1s
Straight / Diverging	56.9s	52.7s	52.7s

Table 11-4: Limiting Operational Headways for Diverging Junction

If the results of Table 11-4 are compared with those of an in-line station stop, it can be seen that diverging junctions are also able to offer far higher capacities than could be utilised with an in-line station stop. This result is the same as that for the higher speed main line scenario.

11.5 MULTIPLE PLATFORM STATION HEADWAY

If a second off-line platform is added to a station area, the scenario becomes as shown in Figure 11-5. It should be noted that on a typical metro railway, all trains stop at all stations. As a result, there are only two route options for any train:

1. Stopping in the in-line platform and;
2. Stopping in the off-line platform.

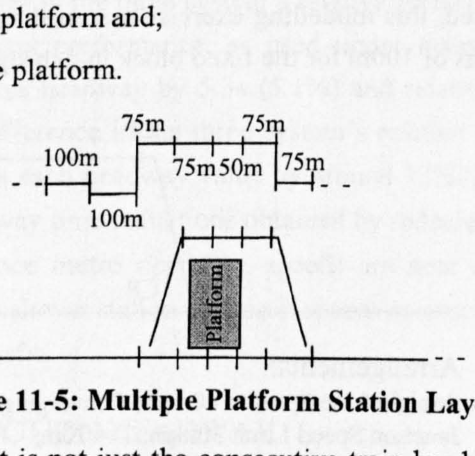


Figure 11-5: Multiple Platform Station Layout

As with the main line case, it is not just the consecutive train headway that is of interest, but also the headway between consecutive trains to the same platform.

The results of implementing the multiple platform station model for the arrangements represented in Figure 11-5, with all trains stopping, are given in Table 11-5. The main figures are for consecutive trains following the pattern indicated in the left hand column. The supplementary figures (in brackets) indicate the headway required between consecutive trains to the same platform with a train to the other platform in between.

Scenario	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
Straight / Straight	95.3s (121.3s)	90.0s (112.9s)	85.8s (112.9s)
Straight / Diverging	60.3s	56.1s	56.1s
Diverging / Diverging	89.4s (121.3s)	88.8s (112.9s)	83.8s (112.9s)
Diverging / Straight	61.0s	56.8s	56.8s

Table 11-5: Limiting Operational Headways For Two Platform Station With All Trains Stopping In Alternate Platforms

The results of Table 11-5 show that, in the 30s dwell case considered, platform occupancy is not a constraint on throughput. Overall, a more significant benefit is gained from introducing moving block than was obtained at the higher main line speeds (the improvement being around 7% in this case, where it was only 1% at higher speeds). In both scenarios no additional benefits can be gained by use of relative braking.

11.6 COMPLEX SCENARIO COMBINATIONS

As with main line railways, most metros would in fact consist of a combination of the basic scenarios that have been considered so far in this chapter. Therefore, it is necessary to consider the overall effect of the various ATP systems when applied to realistic railway layouts. To this end, the author considered two particular complex scenarios to be of interest in determining the overall impact of different ATP types on a metro railway:

1. Traffic from two lines merging at a converging junction, all trains then stopping at a single in-line platform station, before splitting on to two lines at a diverging junction;
2. Repeating the same complex scenario, but replacing the single in-line station with a two-platform station.

These scenarios can be built from the discrete scenarios, as shown in Table 11-6.

Route Stage (all stops dwell for 30s)	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
30mph Converging Junction conv / str (str / conv)	44.9s (51.2s)	36.0s (42.4s)	26.2s (32.6s)
Plain Line, 40mph PSR	42.1s	30.9s	23.1s
Single In-line Station Stop	95.3s	90.0s	85.8s
2 Platform Station Stop, all stop. Div/str (str/div)	61.0s (60.3s)	56.8s (56.1s)	56.8s (56.1s)
30mph Diverging Junction div / str (str / div),	60.6s (56.9s)	56.4s (52.7s)	56.4s (52.7s)

Table 11-6: Limiting Operational Headways for Complex Arrangements

In every case the single in-line station stop is by far the most restrictive element in terms of capacity. Even if all stations have two platforms, the most significant scenario element is still the station stop in all cases. As a result, on a metro line it is always the station areas that determine the overall headway and the impact of ATP types on steady state capacity.

For comparative purposes, the author extracted the critical headway times from Table 11-6 and determined the percentage capacity change that would be obtained by changing from fixed block to moving block or relative braking. The results of this comparison can be seen in Table 11-7.

ATP Arrangement	Limiting Headway	% change on fixed block
Single In-line Station		
Continuous In-Cab ATP (fixed block)	95.3s	-
Moving Block	90.0s	5.6% less
Relative Braking	85.8s	10.0% less
Two Platforms Per-direction at all Stations (average headway values)		
Continuous In-Cab ATP (fixed block)	60.7s	-
Moving Block	56.5s	6.9% less
Relative Braking	56.5s	6.9% less

Table 11-7: Overall Impact of ATP Types on Capacity with a Single In-line Station Stop (all trains stop)

In the case of a typical metro railway with at least one in-line station stop included within the route, the benefit to be gained by changing from fixed to moving block is only about half that which was predicted for higher speed main line railways, mainly due to the improved optimisation of fixed block track circuit locations. However, the further benefit to be gained by changing to relative braking is slightly more significant.

If the line were to have two platforms per direction at every station, an unlikely occurrence on metro railways, the benefit of moving block would again be more significant than in the main line case, whilst relative braking would provide no additional benefit.

11.7 TARGETED MOVING BLOCK

Analysis of the results already presented in this chapter reveals that the targeted application of moving block would offer potential benefit to a metro, but not such significant benefit as found in the main line case (see Table 11-7 and Table 10-16). Localised application to single in-line platforms would in fact achieve only a small benefit of 5.6%. Since the risks associated with implementing a localised moving block would (as a novel system) in the first implementation be higher than those for a full moving block application, the case for targeted moving block is not so strong on a metro. Whether it would be worth perusing at all would again depend on:

- Whether the cost and complexity of localised moving block would be significantly different to that of a full moving block system;
- Whether any potential cost / complexity savings would be worth the loss of full moving block's inherently improved perturbations response and;
- The level of increased risk due to system novelty.

The answers to these questions again fall outside of the scope of this study, but would certainly warrant further investigation.

11.8 SENSITIVITY TO BRAKING CHARACTERISTICS

At low speeds, the majority of the headway distance is not determined by the braking distance, but by the fixed allowances for safety margins, train length and sighting / processing times. Although increasing operating speeds in response to improved service brake rates still increase the braking distance proportionally with the square of the speed, the increased running time over this extra distance is less significant than the corresponding time savings over the fixed margins. This result in a positive overall impact on headway time and means that the optimal strategy for braking on low speed metro lines will not necessarily be the same as that for high-speed main lines.

An interesting comparison with the high-speed results of Table 10-17 can be obtained by considering lower speed operations. Varying the service and emergency braking rates for an in-line station stop with a 40mph line speed gives the results of Table 11-8 and Figure 11-7. In this case, a capacity improvement of up to 11% is obtained by increasing both the emergency and service brake rates, from those assumed in the rest of this chapter as typical of current UK metro brake rates, to those achieved on some equivalent railways elsewhere in the world (see chapter 5). Increasing the emergency brake rate alone also provides capacity benefit, but this is limited to 5.1%. As with the main line example already considered, run-time remains dependent on the service brake rate alone. The optimal solution for both maximising capacity and minimising run time on a metro railway is to increase both service and emergency brake rates.

Emergency Brake Rate	Service Brake Rate	Moving Block Headway	Run-time over 2500m scenario
$1.15m/s^2$	$1.04m/s^2$	90.0s (40tph)	187.4s
$1.5m/s^2$	$1.04m/s^2$	88.1s (40tph)	187.4s
$1.5m/s^2$	$1.4m/s^2$	85.9s (41tph)	185.1s
$2.8m/s^2$	$1.04m/s^2$	85.4s (42tph)	187.4s
$2.8m/s^2$	$1.4m/s^2$	83.1s (43tph)	185.1s
$2.8m/s^2$	$2.8m/s^2$	79.9s (45tph)	181.9s

Table 11-8: Brake Rate impact on Headway & Run Times for 40mph line speed

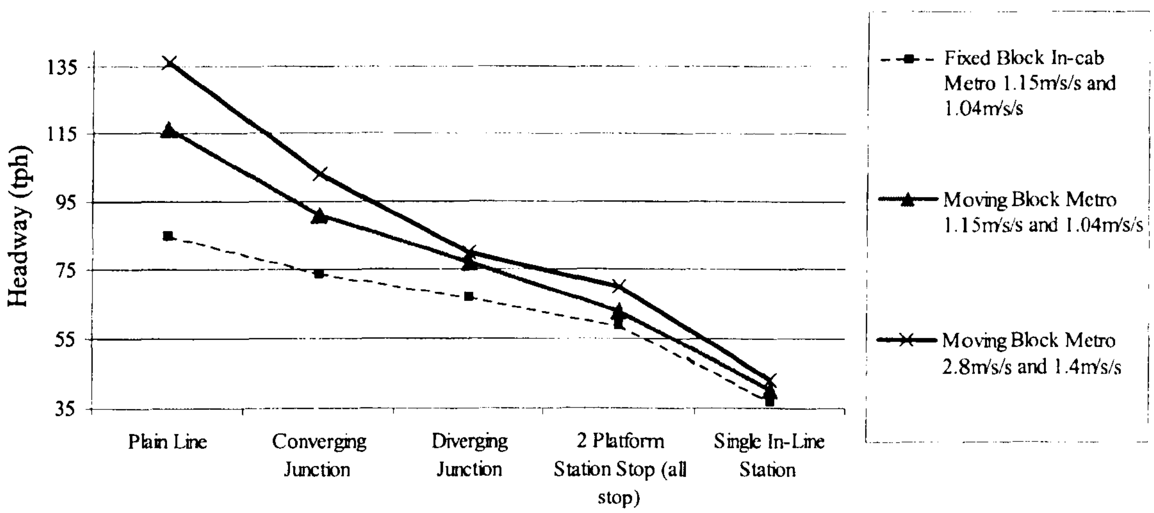


Figure 11-6: Multiple Platform Station Layout

These results can be explained by reference to Figure 11-7. Here the high $1.5m/s^2$ brake rate curve would have a headway time of 29.4s (run in) + 30s (dwell) + 17s (run out to clearance point). This is a total of 76.4s, assuming a 100m train length, 25m safety margin, 20m position error & $1m/s^2$ acceleration rate. In contrast to this, a $1.04m/s^2$ deceleration rate on approach to the station would give 32s + 30s + 17s = 79s. i.e., 2.6s higher in this specific case (note that the ATP sighting point does not change because the train has not yet begun to break to rest in the station when it is reached).

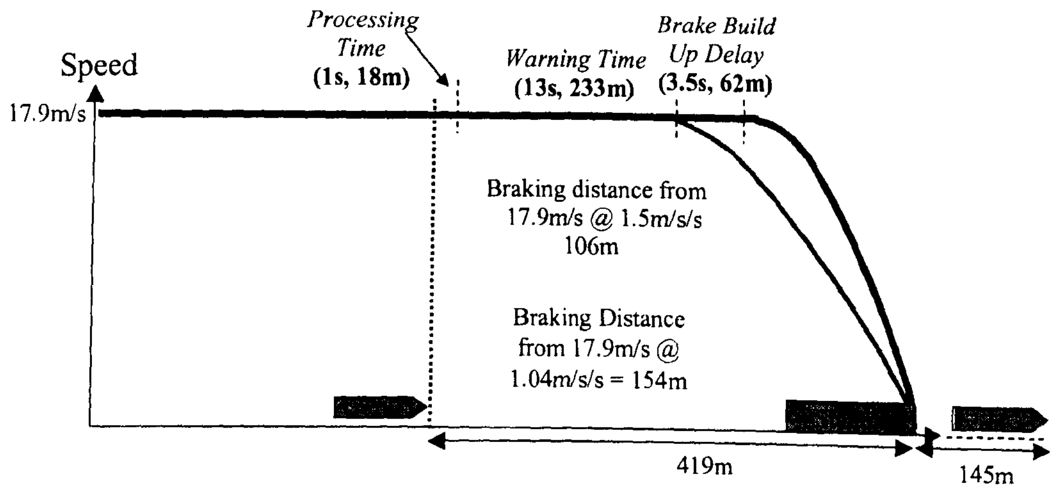


Figure 11-7: Headway Impact of Braking Rates on Low Speed Station Stop Approach

11.9 SUMMARY

The analysis of the elemental scenarios within this chapter has revealed the theoretical capacities of each ATP type, enabling comparison between them in individual scenarios. A summary of the findings of this analysis can be seen in Table 11-9 and Figure 11-8.

Elemental Scenario	Continuous In-Cab (fixed block)	Moving Block	Relative Braking
Plain Line, 40mph PSR	85tph	116tph	155tph
Single In-Line Station Stop (30s dwell)	37tph	40tph	41tph
30mphConverging Junction (average)	74tph	91tph	122tph
30mph Diverging Junction (average)	67tph	77tph	82tph
Multiple Platform Station (alternate platforms)	59tph	63tph	63tph

Table 11-9: Operational ATP System Throughput by Elemental Scenarios

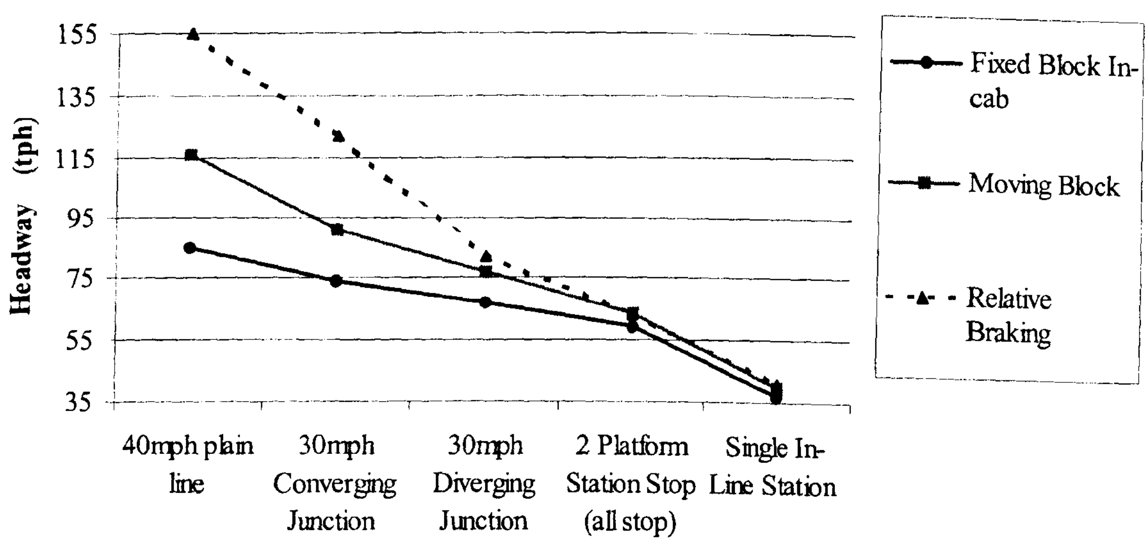


Figure 11-8: Operational ATP System Throughput by Elemental Scenarios

Once again, it is in the complex scenario combinations that the true impact of different ATP types on an overall line's performance is to be seen. As already discussed in section 11.6 a single in-line station stop, if present, is always the limiting elemental scenario for capacity. In such situations, moving block would provide an increase in capacity over fixed block of 5.6%

and relative braking of 10%. This potential benefit for steady state capacity improvement from moving block and relative braking could also be obtained by limited application to the station areas only, with the rest of the line fitted for fixed block in-cab operation.

If all stations include two platforms per direction, the use of moving block or relative braking would be expected to impact capacity by 6.9% when compared with the use of a fixed block in-cab system. In such a case, the application of localised moving block would appear to offer insufficient benefit to warrant implementation, whilst relative braking would offer no benefit at all.

Overall, the benefit of moving block would appear to be marginal and certainly far less than predicted for higher main line speeds. However, it is the author's opinion that, in light of the already proven technology and the demand for increased capacity on metro railways, the benefit is still sufficient to justify the selection of a moving block, rather than a fixed block, system.

Whilst the additional capacity benefits of relative braking are in the metro case roughly comparable to those of moving over fixed block, they would appear to be too limited to warrant the additional risk that such a system would inherently bring to the railway.

12 FINDINGS AND CONCLUSIONS

12.1 INTRODUCTION

In the preceding chapters of this report, the characteristics of railways and their control systems have been considered. In the process of doing this, the potential has been shown for new approaches to train control that could address both the technological need for change and the end user pressure for improvements to railway services. As a part of this, the primary purposes of train control systems were identified in Section 2.1 as being:

- To maintain safety

and

- To provide and manage operational capacity

At the core of these two objectives is the concept of automatic train protection, the stated subject of this thesis, which has been identified as offering the greatest potential of all recent technological developments for improving railway safety, along with significant potential to impact railway capacity. However, whilst acknowledging this, it is important to note that the historical development of 'signalling' based train control was hampered by the splitting of train control functions into a number of technical sub-systems managed by several operational departments (see section 3.12).

The focus of attention must not be fixed on any individual aspect of the railway system if this mistake of the past is not to be replicated in future train control system development. Instead it must remain open to consideration of the interactions that exist between subsystems and disciplines. As a consequence of this, whilst the author's research has focused on optimisation of ATP systems, other related systems (especially around the train / track interface) have also been considered where the optimisation of ATP alone would clearly not enable the full potential of the railway system to be achieved.

The extensive literature review covered by chapters 2 to 4 of this report included a number of significant findings concerning train control systems and the field of ATP in particular. Together with the interviews with railway professionals conducted by the author, as detailed in appendix A, these findings pointed towards a shortlist of areas to consider for the further optimisation of ATP systems:

- Effects on safety performance;
- Cost (of design, installation and maintenance);
- Effects on capacity (achievable headway, availability).

12.2 SAFETY

Whilst safety must always be a primary concern when considering ATP system, no significant scope for enhancing the safety of ATP systems was identified during the literature review. However, it was noted that ATP systems have not been widely implemented within the UK. A clear opportunity for enhancing railway safety therefore exists through their wider installation and use.

The analysis in chapter 2 led to the conclusion that despite public demand for the installation of safer control systems (such as ATP) in the UK, such systems are not commercially justifiable on the grounds of their improved safety alone. This conclusion has also been drawn by the UK based National ERTMS Programme, particularly in the light of TPWS implementation making the safety benefits of more comprehensive ATP systems almost negligible (Informed Sources 2004, p24). In consequence of this, the author did not consider 'Safety' to warrant specific research in its own right during his studies. None the less, some important conclusions about railway safety arose from the author's literature review.

12.2.1 THE INTRODUCTION OF RISK TO AN OPERATIONAL RAILWAY (CHAPTER 2)

The discussion of railway risk within Chapter 2 was to consideration of: ALARP; the limits of tolerability; public perception and user demand – concepts that underpin the approach taken towards safety on today's railways in the UK. Following the approach that these concepts exemplify, there is a clear need for rail safety to continuously improve by the introduction of systems such as ATP.

The same chapter also included discussion of financial appraisal as applied to railway projects, outlining the current trend towards cost benefit analysis. This allows for wider project benefits to be included when justifying expenditure on a railway project (such as the socio-economic impacts of changes in capacity, safety and reliability – that is, economic impacts beyond the scope of the railway system itself).

Building on these discussions, the author wrote a paper that was published in the IRSE proceedings, proposing that wider socio-safety impacts should be considered within the safety evaluation of railway projects (Woodland 2001). He argued that this would allow the adjustment of railway project risk tolerability limits, making it possible to introduce additional risks to the railway (effectively increasing the risk to rail users) if in so doing overall transportation risks were reduced.

An example to illustrate the possibilities of this approach can be readily found in the aftermath of the October 2000 Hatfield derailment. In order to reduce the risks of a similar accident occurring again before other worn sections of rail could be replaced, speed restrictions were imposed across the network. This had the effect of making many journeys longer than they had been in Victorian times, if the train was not cancelled altogether. As a result, rail passenger numbers fell significantly as they abandoned rail in favour of road travel. This not only caused considerable traffic congestion and delays to road journeys, but also significantly increased the risk to travellers (since the accident fatality rate per kilometre is 12 times higher on roads than on railways). It was estimated that by November 2000 the growth in road traffic due to this short term application of railway speed restrictions had resulted in five extra deaths. In comparison to this, broken rails had caused 6 fatalities in the preceding 30 years, including the 4 at Hatfield (The Economist 2000, p2). Efforts to reduce railway risks had therefore resulted in societal dis-benefits and significantly increased the overall risk experienced by the travelling public.

In principle, the author believes that the inclusion of socio-safety in railway project risk analysis is a sound approach that could enable the development of cheaper rail systems with higher capacity and reliability. This would in turn attract passengers away from road transport. Whilst these systems would have lower safety levels than existing railways, so long as their safety levels remained substantially higher than those achieved on the roads, their increased usage would still increase overall transportation safety. Here, however, the most serious problem with the concept of introducing risk to one area of transportation (railways) whilst decreasing overall transport risk becomes apparent. This is the problem of responsibility. Within the UK legal system, and following the guidelines of the ALARP process, risks within the railway system are firmly the responsibility of the railway industry, whilst risks within other transport modes are not. Hence, it was acceptable after Hatfield for the railways to reduce their own risk levels at the expense of increasing the wider transport system risks. Conventional thinking thus makes increasing railway risks in order to reduce overall transportation risks unacceptable. To do otherwise would leave the railway authorities exposed to accusations of failing to meet their safety obligations towards passengers. Hence the laudable aims of the legal requirements and the ALARP principle actually prevent the achievement of optimal safety for society as a whole.

In the light of the current legal framework, it would appear that any decision to implement an optimal re-distribution of risks between transport modes would have to originate from decisions by the authorities responsible for the safety of all modes - ultimately a political decision by government, supported and enforced by all relevant regulatory bodies and the law. Without active involvement at this level it will not be possible for any railway authority to sanction such a change in approach, as they would be unable to do so whilst still demonstrating the risks within their control to be ALARP.

This view would seem to be supported by reports in the aftermath of the Hatfield derailment, which stated that "Railtrack officials privately admit that if they could be sure of political support, they would lift many of the speed restrictions immediately" (The Economist 2000, p3). In effect, the lifting of the speed restrictions before completion of remedial works would have been an introduction of risk to the railway that did not exist before the act of lifting the restrictions. However, it was felt that there was inadequate political support, despite political pressure to return the service to normal. The restrictions were therefore maintained until the remedial works had been completed.

It is interesting to note at this point that it would currently be acceptable to increase one aspect of a railways risk on the basis that overall railway risk has been maintained or reduced by other aspects of the project. It would also be acceptable to increase overall railway risk in the short term in order to achieve a "sustained decrease in risk in the longer term" (Railtrack 2000, p8-4).

It is the conclusion of the author that the political will does not currently exist to enable risk optimisation across the transportation system as a whole, but that, in the interests of society, efforts should be made to develop that will. In the mean time, by judicious use of the safety assessment and financial appraisal techniques discussed within this thesis, the means already exist to justify the introduction of some new risks to elements of the railway system where the overall effect on safety on the railway as a whole can be shown to maintain or improve on the status quo.

12.2.2 HUMAN FACTORS (CHAPTER 4)

It was noted in section 4.3.8 that the introduction of automation often leaves the remaining human operators with more complicated tasks that they must carry out with a reduced sense of purpose and understanding of what is happening within the system, making the occurrence of poor quality work and human error all the more probable. Therefore, an overall systems approach must be adopted when implementing automation, in order to allow for the effects of technical failures and interactions with staff, passengers and the public.

In general, automation works best where humans and machines are integrated within the system such that the machines assist, rather than replace, the skills and abilities of human operators, minimising their workload whilst maximising the use of their decision-making skills.

Unfortunately, "it is absolutely clear that the railway industry, including signal engineers, do not understand enough about human factors and cannot demonstrate how they have been taken into account in the development of railway systems" (Cooksey 2001, p4). This is a weakness that must be addressed in the development of future systems for the automation of railway control.

12.2.3 RAILWAY 'SYSTEM' THINKING (CHAPTER 2)

The author found that optimal solutions for enhancing railway performance may not always lie in the area of train control, nor are they always best solved by the application of advanced technology. The largest risks on railways are, for example, not usually directly associated with train control, but are more likely to include issues such as vandalism and trespass, along with more general health and safety issues like slips, trips and falls. For optimal safety performance it may be better to spend available money on mitigating these risks rather than making train control safer.

Instead of striving to prevent train collisions by implementing ATP, perhaps the industry should be seeking ways in which control systems can assist in the reduction of the more significant (and mundane) risks. Not only would such an approach improve overall railway system safety, it may also provide a means of justifying control system solutions, including ATP and ATO, that would otherwise be difficult to justify. For example, with some effort to understand overall system requirements and apply cross-discipline solutions, the intruder and obstacle detection systems required for driverless ATO could also be used to help in reducing the risks associated with vandalism and trespass.

12.3 COST (CHAPTER 2)

The issue of cost is more complicated to address than that of safety. The concept of 'cost of a life' has already been considered in section 2.4.2 and cost/benefit analysis in section 2.7. It was also shown in section 2.4.3 that, during the 32 years between 1967 and 1999 (representing a generous time scale for the life of any of modern microprocessor based system), there were 31 major ATP preventable accidents, resulting in 122 fatalities. On the basis of these statistics it was further shown that consideration of 'cost of a life' alone would have justified the expenditure of £244 million on ATP systems to prevent these fatalities, whilst a more thorough analysis (allowing for other accident costs as well) could have been used to justify a far higher expenditure of £775 million.

Whilst these figures sound high, they are put into perspective if the actual costs of a comprehensive ATP system are considered. To this end, it has been estimated that expansion of the BR ATP schemes throughout the UK network, excluding freight only lines, sidings and shunting vehicles, would cost in the region of £1153 million for both track infrastructure costs and train fitment (Davies 2000, p80). If the introduction of ATP is to be financially justifiable, it is therefore clear that either:

i. The initial costs of the system must be reduced significantly or;

ii. Additional benefits, such as increased capacity or reductions in whole life costs, must be sufficient to make up the shortfall in the cost/benefit analysis.

During the literature review and interviews conducted by the author, cost was repeatedly found to be a concern, but only one of secondary importance. It is particularly worth noting that:

- An attempt to achieve significant reductions in initial system costs has been made in the UK, in the form of the 'partial ATP' solution of TPWS (see Appendix C). This is expected to prevent 70% of all ATP preventable accidents and 59% of ATP preventable fatalities (Evans[(2)] 2000, pp 112 &114). However, the costs associated with this system are still expected to exceed £400 million (see interview with D Fenner, Appendix A). This makes the cost / benefit ratio similar to that offered by BR ATP, whilst leaving a significant proportion of potentially preventable accidents unprotected. Therefore, the attempt has not been successful;
- Similar attempts to reduce ATP system costs by reducing the benefits offered are likely to be unacceptable in the future as comprehensive ATP systems, such as the ERTMS systems, become generally available. Indeed, there have been high profile calls for the implementation of ERTMS to be made a legal requirement on all lines (Uff et al. 2001, p120 - findings of the Southall and Ladbroke Grove joint enquiry);
- Additional benefits, such as capacity improvements, are widely seen as a far more important enabler for the wider and quicker application of comprehensive ATP systems than reduction in capital expenditure or whole life costs (see Appendix A).

When considering these points, it was again the author's opinion that there is not currently sufficient justification to warrant research specifically into cheaper forms of ATP.

'Capacity' is thus left as the primary area for further research, with cost and safety as secondary issues to be considered when evaluating options for improving capacity through optimisation of ATP systems.

12.4 CAPACITY

The capacity of railway systems is determined by a complex combination of factors, as outlined in section 2.5. A number of these factors relate to the type of train control systems used and the way in which they are implemented. The literature review and interviews conducted by the author allowed him to identify the main areas of interest for the optimisation of ATP and related systems as the relative capacity impacts of:

- Train separation strategies, including fixed block, moving block and relative braking, and
- Data transmission methods, such as intermittent colour light signals, intermittent balises or continuous radio transmission.

In addition to this, he identified two related factors that could have significant impacts on the capacity of railways fitted with ATP systems:

- Train braking performance and;
- Speed supervision criteria.

These four areas of interest then formed the basis of the author's detailed research. The main conclusions drawn from this research are outlined in the remainder of this chapter.

12.4.1 TRAIN BRAKING PERFORMANCE (CHAPTERS 5, 10 AND 11)

As reported in Chapter 5, the major limitation that applies to maximum achievable braking rates today is the adhesion available between the rail and wheel, and its variability in particular. The impact of this limitation can be reduced by means of existing techniques for using available adhesion (such as WSP) and by a comprehensive approach to adhesion management (including vegetation control, sanding, chemical augmentation, wheel/rail cleaning and track maintenance). They can also be reduced by adopting existing technology for braking systems independent of wheel/rail adhesion.

Based on a review of existing braking system implementations throughout the world (predominantly Japan, Germany and France), significant improvements could be obtained in UK brake rates within the constraints imposed by adhesion and passenger physiology, equivalent to:

- 60% (from $0.88m/s^2$ to $1.4m/s^2$) for main line service braking;
- 28% (from $1.17m/s^2$ to $1.5m/s^2$) for main line emergency braking;
- 26% (from $1.15m/s^2$ to $1.45m/s^2$) for heavy metro service braking and;
- 115% (from $1.3m/s^2$ to $2.8m/s^2$) for heavy metro emergency braking.

Modelling of the effects to be expected from application of such improvements to a moving block ATP system has been documented in chapters 10 and 11. In summary, it was found that:

- At a 125mph line speed (typical for UK main line railways) the maximum capacity would be achieved with a high ATP intervention brake rate and relatively low service brake rate (as applied by the driver) for stopping at stations or signals. Taking the current values typical on UK main line railways to be $0.88m/s^2$ and $0.49m/s^2$ respectively, increasing the ATP intervention rate to $1.5m/s^2$ whilst retaining the same driver applied rate would give a headway time improvement of 17%, increasing the throughput from 21 to 25tph;
- Increasing the driver's applied brake rate at a 125mph line speed to $1.4m/s^2$, in conjunction with an ATP intervention brake rate of $1.5m/s^2$, would actually increase the headway time by 1.4%, but would reduce the run time over a 12km scenario (including a station stop) by 9.4%;
- At the lower 40mph line speeds typical of metro services, the maximum capacity will be achieved when both the ATP intervention brake rate and driver's service brake rate for stopping at stations or signals are as high as possible. Taking the current values typical on LUL to be $1.15m/s^2$ and $1.04m/s^2$ respectively, increasing the ATP intervention rate to $2.8m/s^2$ and the driver's application rate to $1.4m/s^2$ would give a headway time improvement of 8%, increasing the throughput from 40 to 43tph, whilst also reducing run time over a 2.5km scenario that includes a station stop by 1.2%.

For either railway type, implementing such increases in brake rate clearly offer significant potential for capacity improvement.

Unfortunately, significant and sustainable enhancements to train braking can only be achieved by introducing new stock. Due to the costs and timescales involved in rolling stock production and acceptance, this would necessitate a gradual introduction rather than a big bang approach - making optimisation of any fixed block system impossible until all trains had been upgraded.

Whilst this may be acceptable for a metro railway where whole fleets are often replaced in a short time frame, it is unlikely to be so for main line railways. In contrast, the enhanced performance characteristics of any upgraded train could be utilised as soon as it began operating if moving block control is provided. Some capacity improvement could then be expected from commencement of the rollout. This is a potential benefit of moving block ATP that should be considered when determining an optimal strategy for ATP implementation.

12.4.2 SPEED SUPERVISION CRITERIA (CHAPTER 6)

Allowing signed permitted speeds to be increased for ATP fitted trains in general was not found to be practicable in chapter 6, because the main limiting factor in speed setting is passenger comfort, not safety. However, the potential for raising some temporary speed restrictions and / or tilting train enhanced permissible speeds was identified. This could be done by fine-tuning ATP supervision criteria in order to reduce probable over speed where the predominant factor in defining the acceptable speed limit is overturning or derailment risk.

In accordance with this objective, a number of areas offering scope for future improvements in ATP supervision criteria were identified. In particular:

- The use of higher braking rates on ATP intervention (reinforcing the findings of chapter 5);
- More gradual control of acceleration by limiting the traction system's response to a driver's demand (particularly at low speeds) as the train approaches the permissible speed;
- The complete cutting of traction power following an overspeed warning;
- The use of speed measurement systems with improved accuracy;
- Elimination of warnings prior to brake intervention, thus enabling reduced margins between permitted and intervention speeds.

Whilst some benefit could be gained by implementing these improvements, the extent of those benefits is not expected to be large.

12.4.3 OPERATIONAL IMPACT OF TRAIN SEPARATION STRATEGIES AND DATA TRANSMISSION METHODS (CHAPTERS 10 AND 11)

The analysis of element scenarios within chapters 10 and 11 revealed the theoretical capacities of each ATP type, enabling comparison between them. This comparison revealed the general trends that:

- ATP overlay systems reduce capacity, with the impact reducing as the update frequency is increased;
- In-cab systems increase capacity, the impact increasing from fixed to moving block and from moving block to relative braking.

When complex scenario combinations were considered it became clear that a single in-line station stop is always the limiting element scenario for capacity. Where one is present:

- For a main line railway (125mph line speed), ATP overlay systems would not be expected to affect the steady state capacity achieved by use of conventional 4-aspect signalling; fixed block in-cab systems would provide a capacity increase of 11 to 15%;

moving block would provide an increase in the region of 29 to 34% (that is, a 21 to 22% increase on fixed block in-cab) and relative braking would add 31 to 36% (a 23 to 24% increase on fixed block in-cab), depending on original 4-aspect signal spacing;

- For a metro railway (40mph line speed) moving block would provide an increase in capacity over fixed block in-cab of 5.6% and relative braking would add 10%. Conventional lineside signalling and ATP overlay systems were not analysed in this case.

In both cases, the moving block and relative braking steady state capacity improvement could be obtained by limited application to the station areas only, with the rest of the line fitted for fixed block in-cab operation.

If <u>all</u> stations include two platforms per direction:

- For a main line railway, the use of ATP overlay systems would be expected to reduce capacity by between 0.3 and 21%, depending on the original 4-aspect signal spacing and the frequency of supervision criteria updates. A fixed block in-cab system would be expected to give an increase in capacity of 19 to 26% and both moving block and relative braking would add 21 to 28% (again depending on the original 4-aspect signal spacing). The increase in capacity achieved by moving from fixed block in-cab to moving block ATP would therefore be in the region of 1%;
- For a metro railway, the use of moving block or relative braking would be expected to increase capacity by 6.9% when compared with the use of a fixed block in-cab system.

In such a case, the application of localised moving block in station areas would appear to offer insufficient benefit to warrant implementation, at 1.2 to 1.3%, depending on stopping patterns. Localised use of relative braking in the station area would offer no benefit at all.

Overall, the findings (summarised in Figure 12-1) suggest that the use of ATP overlay systems would only be appropriate on a line that is lightly used when compared with the design capacity of its existing lineside signals. On lines operating at or near their lineside signalling design capacity, the case for overlay ATP implementation is likely to be undermined by the capacity reduction that would be imposed. Similarly, the total cost of lineside signalling replacement and overlay ATP implementation on lines requiring resignalling is likely to be significantly higher than that of replacement by an in-cab signalling system, again undermining the case for their use.

The results for both main line and metro railway models demonstrate a good case for the implementation of ATP in a continuous in-cab form and for the adoption of moving block rather than fixed block train separations, at least in the area of in-line station platforms. Whilst the benefit to be gained from moving block application is higher on main lines than it is on metros, the demand for increased metro capacity on many of the worlds metro railways is sufficient to make the smaller benefit remain highly attractive – particularly since the viability of moving block has been proved by several equipment manufacturers on metro railways throughout the world.

By contrast, introducing a significant increase in collision risk as a part of implementing unproven technology would appear to undermine the case for developing moving block into relative braking train separation.

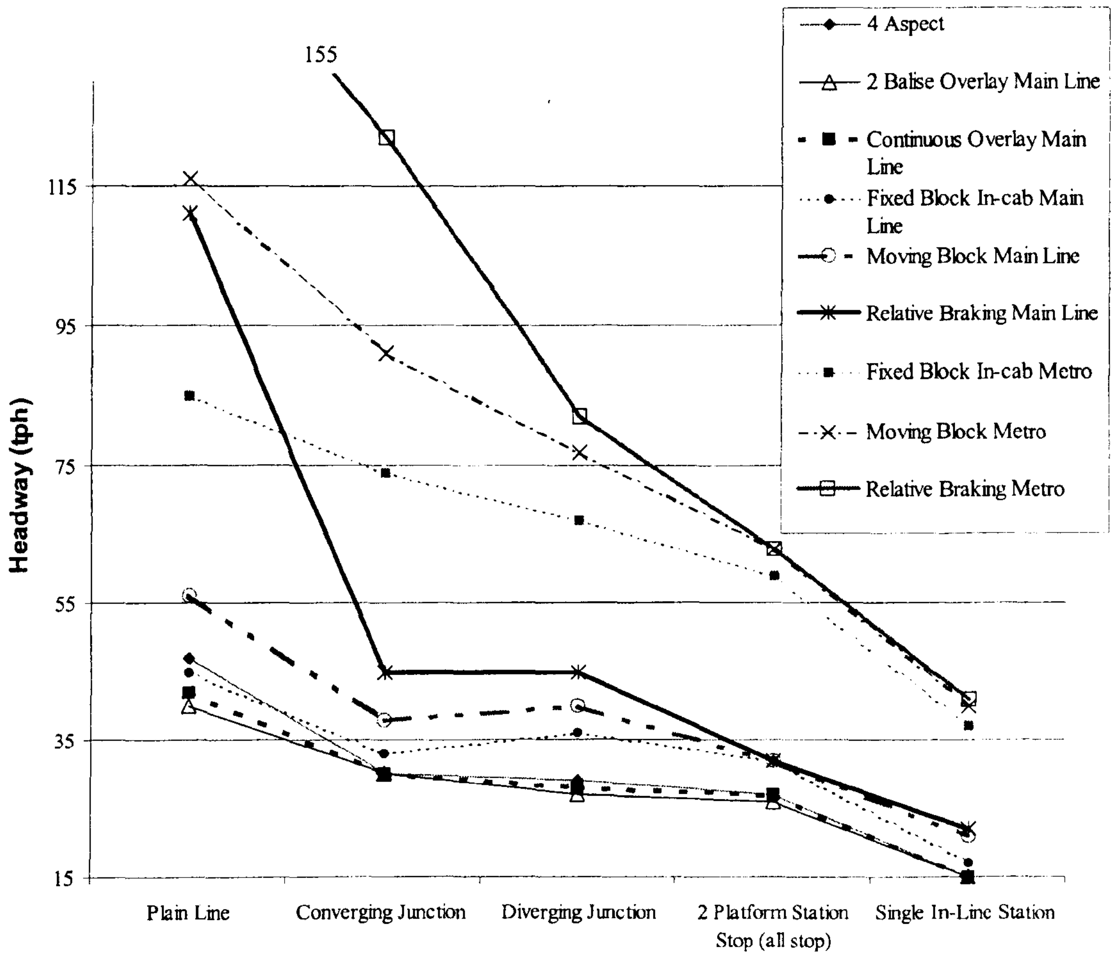


Figure 12-1: Operational ATP System Throughput by Elemental Scenarios

12.4.4 RAILWAY 'SYSTEM' THINKING (CHAPTERS 2, 5 AND 6)

As already noted in the discussion of safety, optimal solutions for railway performance enhancement may extend beyond train control and may not require the use of advanced technology.

The potential for adopting enhanced train braking performance has already been mentioned in section 12.4.1. Alternative, less high-technology, train based solutions for capacity increase can also be found through initiatives such as increasing train length, providing additional doors and modifying seating arrangements.

If station control is considered, improved platform management can gain significant capacity benefits. Practical initiatives on LUL's Victoria Line have demonstrated dwell time reductions of around 10% through improved platform management, based on platform staff activities and without the need for any significant additional equipment (Horsey 2000, pp2, 73).

The railway industry as a whole needs to put far more effort into understanding cross-system (multi-discipline), cross-network and even cross-modal issues and solutions if it is to develop a rail transport product that provides an optimal contribution to society's transportation needs. The author hopes that the multi-discipline discussions included in chapters 2, 5 and 6 of this thesis will stimulate some further development in this area.

12.5 ATP SYSTEM ISSUES

Whilst the author's study has focused on capacity issues, a number of wider issues became apparent during his review of relevant literature, some of which raise important questions about whether the approach being adopted within modern ATP system design is really optimal. Some of these questions have not been addressed within this thesis so far, although discussion of them can be found in papers published by the author. However, they form an important part of the conclusions that the author has been able to draw from the overall study.

12.5.1 CONTROL & PROTECTION (CHAPTERS 3, 4 AND APPENDIX C)

In most industries, the functions of control (management of day to day operations) and protection (intervention in case of unsafe control demands) are kept separate – using diverse data, software and hardware. This is done for good reasons:

i. It is advantageous to keep protection systems as simple as possible (and thus reliable as well as predictable) and

ii. It is vital that the protection system continues to function correctly even if the control system fails to do so (and vice versa).

The introduction of modern microprocessor systems has enabled significant integration of functions and inclusion of automatic protection within control systems. In such cases, large cost savings can be gained by avoiding the need for duplication – but these savings bring with them a significant risk of simultaneous control and protection system failure. For this reason, whilst equipment suppliers offer such products for use in chemical and material processing plants, independent protection systems are often retained – even though this incurs additional costs.

In some industries, the argument for integration of control and protection functions is not based on cost alone. A classic example of this is the aerospace industry – which, in the provision of public transport vehicles, offers an interesting comparison to railways. In the case of passenger aircraft flight control systems, mechanical linkages are now commonly being replaced by digital fly-by-wire control systems. These enable greater flexibility, improved response to control demands, the possibility of more advanced autopilot systems and the facility to automatically protect against control demands that could overstress the airframe or result in a stall. However, they also offer inherently lower integrity than their mechanical predecessors. The attraction of fly-by-wire control lies more in the improved functionality than in the cost of the system. Indeed, the costs can actually be higher since, in order to enhance the overall system integrity, it becomes necessary to provide redundancy in the form of multiple (usually identical) signal sources, computers and data transmission routes, all with internal and cross channel fault monitoring in order to isolate any failed equipment and ensure safe operation.

A modern civil aircraft has at least two flight management systems, two auto pilots and two auto-throttle controllers, designed with an overall probability of failure less than 10^{-9} per flight hour. However, this alone is not considered to be enough. As an additional safety measure, provision is also made to switch to a dissimilar redundant (usually analogue and/or mechanical) backup system if the digital system fails. Circuits providing warning signals in the cockpit are also usually kept independent of the circuits or systems providing the controlling actions. The dissimilar system may offer less functionality than the primary fly-by-wire control, but this approach guarantees that a flight-critical function is not lost if a generic software, hardware or data fault causes the failure of all identical redundant components (Fielding et al. 2000, pp26, 29-30; Moir et al. 2001, p211).

Unfortunately, whilst advanced digital fly-by-wire systems bring the benefits of improved control and increased integration, they also result in greater levels of complexity in the development process. This can make it far more difficult to prove safety and ultimately obtain safety certification. This fact has not been lost on the aerospace industry. As a consequence, the integration of functions is not always carried out, even when technically possible. The benefits to be gained are instead assessed with respect to system safety and availability objectives, and integration is only pursued where these outweigh the expected difficulties. An example of this can be found in the Airbus A320, where the Air Data Computer and Inertial Reference System were integrated within the same line replaceable unit but retained use of separate computing devices in order to keep the independence of the two functions and ease the certification process (Grossin 1992, p439-441; Moir et al 2001, p289).

On railways, control and protection functions have also been kept separate historically. Ever since the introduction of the interlocking, service control activities (whether carried out by human operators or automated systems) have dealt with control and the interlocking with protection. The control is further designed not to stress the protection system (i.e., it acts as a first layer of safety protection itself, filtering out unsafe route demands before they are passed to the interlocking). Similarly, where there is no ATP system the driver of a train is the primary safety system for train movements. When an ATP system is introduced and the drivers are allowed to continue controlling the train in the same manner as before, they can continue to act as the primary safety system for the train's movements whilst the ATP effectively acts as a secondary safety system (or safety net) in case of driver error.

Early applications of automatic train control, such as the mechanical trainstop, followed this approach, leaving train control with a human operator (the driver) whilst taking protection into an automated system. The human operator would not intentionally rely on the protection system and thus provided a layer of safety independent of the safety critical protection system. This remained the case with the first introduction of passenger service ATO, on LUL's Victoria Line, which used coded track circuit transmissions for train protection and independently coded spot loop transmissions for driving instructions (see Appendix C 7.8). By generating driving instruction codes compliant with both the regulation objectives and safe speeds / train separations, the ATO system effectively retained an independent control and safety layer. However, in more recent times the separation between control and protection functionality has often been lost in train operation.

Whilst the Victoria Line ATO/ATP arrangement ensured independence between control and protection functions, the later Central Line ATO system obtained target speeds from the train borne ATP system (see Appendix C 7.5). Any failure in the ATP system that might result in unsafe target speeds would thus impact both the ATP and ATO, whether caused by hardware, software or data errors.

An equivalent arrangement to the Central Line ATP / ATO systems also exists in most modern ATP systems designed for use with a human driver, including AWS, BR-ATP and ERTMS. As noted in section 4.2.2, if indications of target speed and movement authority are provided to the driver by the ATP system, they will influence the way in which he/she controls the train. Any error in the ATP generated target data is then likely to mislead the driver and cause him/her to make the same error, undermining the independence of control and protection.

The saving grace up to now has been that drivers can observe independent lineside signals and are required by the railway's operating rules to regard them as their primary movement authority. Thus, by following the rules, drivers **may** be able to spot errors in ATP targets and continue to control train movements safely. However, as we begin to move to in-cab signalling systems, with no lineside signals, this will no longer be the case. The ATP system will be providing both protection and control instructions based on the same hardware, software and data, and the driver will have no independent means of assessing the in-cab signal's validity – effectively providing an overall train control system of lower integrity.

Such systems may well still be safe, but proof of that safety is going to become far harder. They could be made safer still, and the safety approval process simplified, by following wider industry best practice and ensuring independence of control and protection functions.

12.5.2 EFFICIENCY, SAFETY AND DRIVING CUES (CHAPTER 2 & APPENDIX C)

It has been noted in chapter 2 that, in fulfilling their principal task of ensuring that trains run safely by maintaining the separation between them, train control systems can also provide a means of regulating the service.

Regulation has in the past been performed very crudely through the signalling system – with trains being either rerouted or delayed by holding signals at red. Regulation in this way is inefficient and can extend the time required to recover from a perturbation. With the introduction of in-cab signalling systems it becomes possible to achieve a much finer level of regulation. However, any attempt to do this whilst in-cab signalling is a part of the ATP system is bound to increase the complexity of the system, making it harder to gain safety approval.

Besides regulation and direct target information, in-cab signalling systems in use today also include some driving cues. An example of this is the AWS bell / horn which, whilst intended as a driver warning system on approach to signals at caution or danger, is also used by drivers to help them assess their location (whether they receive a clear or caution indication), particularly in adverse weather conditions. Whilst such cues may not have been intended for the use to which they are put, drivers do now rely upon them and would need to revise their route cues if they were to be removed. This was one of the main reasons given to retain active use of AWS when the far more comprehensive BR-ATP systems were brought into operation. Although the limited protection provided by AWS became largely redundant, BR-ATP did not provide equivalent driving cues. Incidentally, retention of the AWS warnings also provided an independent (albeit inferior) layer of protection and indication to the driver, boosting the overall integrity of the BR-ATP systems. However, this resulted in drivers having to respond to multiple indications and alarms initiated by the conventional signalling, BR-ATP and AWS – bringing the potential to annoy or even distract some drivers from the safe control of their train.

With ERTMS it is (at least theoretically) possible to provide driving cues through the in-cab display. However, such use further complicates what is predominantly intended as a safety system and may again make safety approval harder to gain.

Clearly, there is potential for in-cab signalling systems to assist the driving task to a significant extent if the benefits of improved signal clarity, enhanced regulation and clear driving cues can be gained without suffering the problems of attention conflict and safety approval. The author believes that this can best be achieved by deciding what the purpose of the displays is! If the purpose were to prevent an ATP intervention, perhaps effort would be better spent designing compatible ATP supervision criteria and operating rules. If it were to provide driving and regulation cues in support of efficient operation of the railway, perhaps the answer is to make the in-cab signalling and regulation instructions independent of the protection system, much as the Victoria Line ATO provides independent driving instructions. With such a system in place, the driver's job could be simplified significantly.

One consistent set of driving guidance could be provided that would help the driver to know what was expected of him/her (rather than what is permitted by the signalling system), including speed restrictions, movement authorities and station stops, etc. An independent and unseen ATP could then protect against errors by the driver or in the driving guidance system's advice. There would be no need for multiple system alarms and indications to distract and annoy the driver, who would instead receive far clearer instructions to support the operational aspects of their driving task. Not only would such an arrangement make life easier for the driver and simplify safety approval, it would also create opportunities for enhanced regulation systems (see Mitchell 2003, p11) – and pave the way for full ATO in years to come.

The pros and cons of such an approach need to be considered further, along with any alternative ideas, if we are to achieve the best from our drivers and regulation systems in the future.

12.5.3 ADHESION MANAGEMENT (CHAPTERS 5 AND 7 TO 9)

It is a core assumption of railway control that trains can stop. Generally, the assumption is valid, but, in conditions of poor adhesion, it may not be so. For example, it was noted in section 5.4.1 that a typical UK main line train would have a nominal full service brake rate of up to 9%g and emergency brake rate of 12%g, yet rail / wheel adhesion levels below 9% are quite common.

Whilst the safety of trains specified to the existing braking performances has proved historically adequate, they have been operated by professional drivers (who know better than to initiate a full brake application in areas where the adhesion is poor) on lines fitted with conventional signalling systems (that usually allow a 25% contingency between the signal spacing and trains theoretical stopping performance). It is not the usual practice to make such allowances in ATP system algorithms. There is certainly no allowance in BR-ATP (apart from the signal's overlap) and only a small 'provision' in ERTMS algorithms (which it is not obligatory to use). Even then, when the ATP intervenes with a brake application, it does so at full brake rate, regardless of the prevailing adhesion conditions. Whilst ATP systems are overlaid on conventional signalling, this should not cause too many difficulties (assuming that the trains are fitted with a reasonable WSP system to mitigate the ATP's high brake rate demand), since the signal spacing allowance still applies. However, as we move to in-cab signalling without lineside signals it is important to reassess the provision for poor adhesion.

Account could be taken of variable adhesion conditions by:

- Retaining the conventional signalling approach of adding a significant margin to the theoretical braking distance, accepting a consequential impact on capacity;
- Developing a better understanding of worst case achievable braking rates, it may be possible to develop an ATP system that supervises against these, rather than nominal rates for dry rail. However, this would again impose a heavy capacity penalty in areas with good adhesion;
- Addressing the problem from a rolling stock, rather than train protection, perspective. Railways in continental Europe and Japan are able to reliably achieve significantly higher brake rates than are currently possible in the UK, even in times of poor adhesion, through combinations of enhanced Wheel Slide Protection, auxiliary tread brakes, sanding and rail / wheel adhesion free braking techniques;

- Disseminating information relating to areas of poor adhesion, perhaps by use of systems such as AEA Technology Rail's low adhesion warning system discussed in section 5.4.1.7, so that drivers and ATP systems can be advised of the need to compensate for poor adhesion or inadequately performing braking systems.

As discussed in chapter 5, significant scope exists for enhancing the braking performance of UK trains. Such enhancement would not only increase railway capacity (as discussed in section 12.4.1), but also underpin existing assumptions on the safe operation of railway traffic. Conversely, continuing towards implementation of in-cab ATP based train control without the development of improved adhesion management techniques is liable to introduce significant risk to railway operation during periods of low adhesion.

12.6 SUMMARY

The conclusions outlined in this chapter cover a broad spectrum of issues related to improving the safety, cost and capacity of railway operation through ATP system optimisation. The key to safety improvement has been shown to be ATP implementation of any form, whilst it has also been shown that safety improvement alone is insufficient to justify the cost of that implementation. Therefore, the key element for future ATP system optimisation is the added value obtained through capacity improvements.

It has been shown (see Figure 12-2) that existing technology can be applied to railway systems to achieve significant benefits, whilst also maintaining acceptable levels of safety performance:

- Enhancement of main line railway capacity by up to 15% through implementing fixed block in-cab ATP systems in place of 4-aspect lineside signalling;
- Enhancement of main line railway capacity by up to 34% and metro capacities by up to 5.6% through implementing moving block train separation;
- Reduction of moving block headway times by up to 17% on main line railways (that is, a reduction of 41 to 45% when compared to conventional 4-aspect lineside signalling with current UK brake rates) by enhancing rail / wheel adhesion utilisation and train braking performance;
- Reduction of moving block headway times by up to 8% on metro railways (that is, 16% when compared to fixed block in-cab ATP with current UK brake rates) by enhancing rail / wheel adhesion utilisation and train braking performance.

The results show clearly that the future must be with in-cab signalling systems if ATP optimisation is to be achieved. The optimal solution for steady state capacity would then appear to lie with moving block, although fixed block in-cab ATP also offers significant capacity improvements over conventional lineside signalling. In order to provide a fuller assessment of the relative benefits of fixed and moving block train separations, further research is now required into their relative performance under perturbed conditions.

In addition to this, it has been shown that the use of localised moving block application in the area of in-line station stops enables the full steady state capacity improvement offered by comprehensive moving block implementation to be achieved. The economics of targeted versus comprehensive moving block application now need to be analysed in order to determine whether localised application is worth pursuing further.

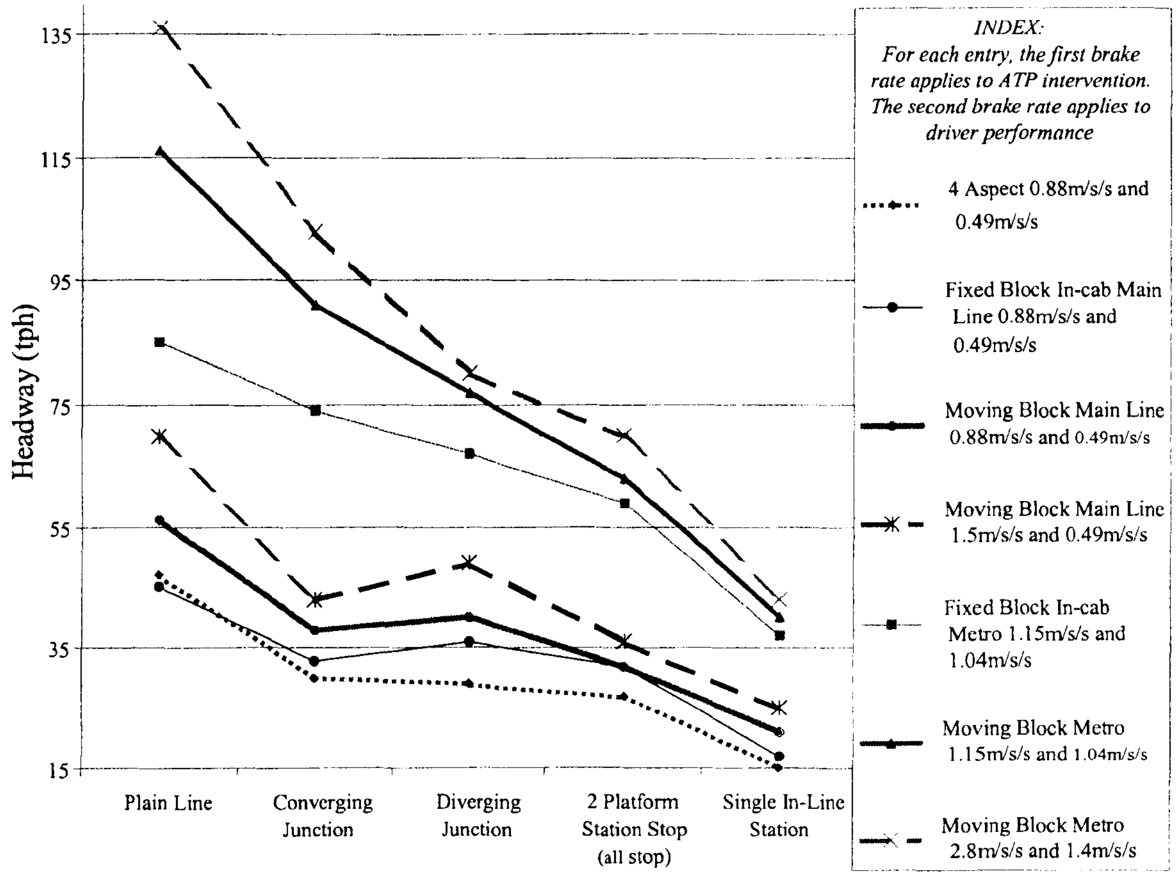


Figure 12-2: Summary of Potential for Steady State Capacity Enhancement

In contrast to the positive results for moving block, there would not seem to be a case for implementing ATP systems with relative braking train separations. Whilst capacity improvements could be achieved, these would be small (up to 2.9% on main lines and 4.6% on metro railways when compared with moving block) and would come at the cost of a significant increase in collision risks that the author believes would outweigh the benefits.

Having made conclusions that appear to be strongly in favour of the introduction of moving block train separation, the author would like to close with a question: In the real world of limited resources, will the interests of rail safety and efficiency always be served best by expenditure on ATP, state of the art moving block control and enhanced train braking?

During the period of study documented within this thesis, the author has concluded that this question can only be answered by further analysis and reference to specific cases:

- New longer trains with more doors, additional platform staff, improved adhesion management or the design of train interiors for better crashworthiness may actually offer better returns on investment for the railway companies and better meet the requirements of their customers;

- Chasing buzzwords such as ‘ATP’, ‘Moving Block’ and ‘In-cab Signalling’ is unlikely in itself to produce the benefits expected unless careful consideration is first given to the overall objectives of the potential enhancement and to the cross-discipline evaluation of all possible means of meeting those objectives;

True optimisation will require multi-discipline cooperation in order to implement improvements in multiple sub-systems of the railway, through coordinated application of technology and appropriate operating practices.

13 FURTHER RESEARCH

13.1 INTRODUCTION

The author's detailed research focused on the impact of ATP systems on railway capacity. Within the timescales available for the research, it was not possible to consider all of the areas of potential interest that were identified. This was particularly true of areas of wider significance for the optimisation of railway control that were identified during his work. There are, therefore, numerous areas to be considered for further research, some of which follow directly from the work of the author, whilst others branch out into related areas.

13.2 DEFINITION OF CAPACITY

Our ability to optimise the capacity of a railway system is dependent, in part, on our ability to define the capacity that we require and to subsequently measure how well we have managed to achieve it. If we are unable to do these two things, we will also be unable to identify the ways in which available capacity can best be used, or to identify the improvements to railway capacity that would satisfy the needs of our customers.

In section 2.5 a general introduction to the subject of railway capacity was given, together with a series of conventional definitions. However, following the literature review and interviews conducted by the author, it is his opinion that this could be taken further. "Customers want a through journey, so the capacity needs also to measure the delivered connection as well as the travel on the train" (Appendix A, interview with D McKeown). It is the author's opinion that it is possible to develop definitions of capacity that focus more on the needs of the customer than the measures considered in section 2.5. If this is done, it may then also become possible to develop ways of measuring the effect of adjusting capacity, not just on train movements, but on the movement of individual passengers (including the effects of waiting times and connection delays).

In pursuit of this aim, the author has identified queuing theory, as used in the analysis of telecommunications and computer network capacity, as a potential tool for enhancing the understanding of railway capacity.

According to queuing theory, a number of factors will affect the capacity of a system:

(a) The arrival pattern (both the average rate of arrival of customers and the statistical pattern of their arrivals);

(b) The service mechanism (when service is available, how many customers can be served at one time and how long the service takes);

(c) The queue-discipline (the method by which a customer is selected for service out of all those waiting for service).

(Cox 1961, p4)

When these factors have been described in sufficient detail for the system being considered, it becomes a mathematical problem to predict what the system will do, including:

- How many people can be expected to be in the system at any given time;
- How long particular queues will be on average;
- The mean and statistical distribution of how long an individual or average customer will spend in a particular queue or the system as a whole;

- The traffic intensity parameter, the units of which are referred to as Erlangs.

Further research into this subject would potentially be of great benefit to the understanding of railway capacity.

13.3 INHERENT SYSTEM DELAYS

The technology used to implement a control system can bring with it inherent delays that will act as a limiting factor on the capacity achievable when using that system. Similarly, the architecture used to implement that technology will have significant effects on system delays. In particular, the factors of interest are:

- Trackside data processing delays;
- Trackside route reset delays;
- Transmission delays between trackside equipment and trains;
- Trainborne processing delays and reaction times;
- Transmission delays between trains and trackside equipment;
- Transmission delays between trains in the same area.

An investigation of these factors was carried out as a part of the author's research and the findings were used to determine the parameters that should be adopted for realistic modelling and assessment during other investigations. However, this investigation focused on what were considered to be realistic delay values for today's systems. No consideration was given to methods of reducing the delays. In the author's opinion, such consideration would be worthy of future research effort.

13.3.1 VEHICLE BASED CONTROL

"Drivers of cars, lorries and buses create a speed dependent virtual space around their vehicle and adapt their speed according to the characteristics of the visible boundary of the space and the speeds and expected behaviour of other traffic partners. Commercial aircraft are allocated slots in corridors and volumes of airspace while operating in congested areas, by air traffic controllers. Like ships they maintain a distance from other planes using radar – in effect an extension of line of sight operation. Private planes use non-controlled airspace by line of sight, if necessary extending the visible horizon by radar. In contrast to all other modes of transport, railways rely on the operation of infrastructure based equipment to set up the correct path at junctions between origin and destination" (Schmid 1999, p1).

One potential means of reducing inherent system delays would be the development of techniques for vehicle-based control on railways, as already used on other modes. Further research on this topic could consider:

- Proposals for Automated Highway Systems (the automation of road vehicle operation);
- Personal Rapid Transit Systems (proposing the use of small vehicles operated on an 'as demanded' basis on guideway systems);
- Current practice for control of air traffic, including arrangements for both commercial and private aircraft;
- Systems already in use, or proposed, for railway systems.

It is anticipated that a review of these topics would provide a useful guide to the potential for, and effectiveness of, vehicle based control in a railway environment. This could then lead on to consideration of the potential for future application of vehicle-based control, including assessment of what functionality could be located on trains, potential mechanisms for achieving this and the impact that it would have on system delays.

13.3.2 DELAYS DUE TO EQUIPMENT RELIABILITY, OPERATION AND MAINTENANCE

The largest factor in non-availability of railway control systems is the availability of wayside equipment (IRSE 1993, p3). Indeed, responsibility for around 40% of all railway delays in the UK is apportioned to the maintenance contractors who are responsible for maintenance, repair, minor renewals and 'rapid response' faulting for the fixed infrastructure contracted to them (Winder 1999 p 2). This is clearly an area where there is scope for improvement, with significant potential benefits:

- Reduced liability for maintenance contractors;
- Reduced delay to the railway's customers;
- Increased actual capacity, bringing it as close as possible to theoretical capacity;
- Increased safety.

The last point is often overlooked and thus particularly worthy of note. "If equipment such as a track circuit fails 'right side' causing a red aspect to be displayed when the signal block is not occupied, the driver of the train is left with an ineffective block working system and must either leave the train stationary or revert to a less sophisticated means of train protection which may itself not have an equivalent level of safety" (Taskin et al. 1995, pA3/10).

In order to achieve these benefits, there are a number of potential areas for further investigation:

- Improved operational rules to minimise the effects of technical failures. "There are doubtless numerous areas of railway operation where the combination of overly restrictive 'rules', and lack of investment in technological improvement combine to impose a heavy penalty in delay, disproportionate to the original failure" (Winder K 1999 ,p4);
- Improved design for fault avoidance. "The first line of defence against failure is fault avoidance, in which design and management techniques attempt to minimise the likelihood of faults arising during specification, development, manufacture and commissioning" (Hazards Forum 1995, p8);
- Improved design for fault tolerance. "The second line of defence is based on the provision of fault tolerance as a means of dynamic protection during system operation". This might include fault masking (where the system or component is designed to survive potential failure with full functionality), graceful degradation (where, in the event of failure, the system is designed to maintain operation, but with the loss of some functionality) and fail-safe design (where, in the event of failure, the system or component automatically reverts to one of a small set of states known to be safe) (Hazards Forum 1995, p8);
- Improved design for fault detection (by use of condition monitoring systems);

- Improved design for maintainability. The squeeze on maintenance access time currently encountered on railways worldwide promotes the need to reduce trackside equipment in order to enable more effective and timely maintenance activities (Winder 1999, p3).

13.4 EFFICIENCY OF IMPLEMENTATION

13.4.1 TRAIN BRAKING PERFORMANCE

A detailed analysis of train braking performance has been documented by the author in chapter 5 of this thesis, including:

- How different types of train braking systems work;
- The braking rates achievable by use of different types of train braking system, either independently or in combination;
- The application delays applicable to different types of train braking system;
- The effects of adhesion and the potential to limit these effects;
- The effect of jerk rates on passenger comfort and safety, including consideration of acceptability criteria;
- Driver utilisation of available braking performance.

The results of the analysis were used to determine the parameters that should be applied for realistic modelling and assessment during other investigations. That modelling in turn demonstrated that significant improvements to capacity could be gained on both main line and metro railways in the UK by adopting best world practice for enhanced braking and adhesion management methods. However, the focus of the analysis was on what could be considered realistic for safe and reliable braking rates today. There is, therefore, a significant opportunity to build on the author's analysis by considering methods by which further enhancement of braking rates could be reliably and safely achieved on railways in the future. In particular, in chapter 5 the author identified three developments that appear to offer significant potential for future braking systems by enabling wider use of wheel/rail adhesion independent braking mechanisms:

- Fail-safe application methods for wheel/rail adhesion independent systems, such as holding a permanent magnet track brake off by use of air or magnets and engaging it by means of springs;
- Control systems capable of adjusting the current applied to an electro-magnet's exciting coils in a fail safe manner, in order to maintain a consistent brake force at all speeds and;
- Localised brake control, so that non-fail-safe mechanisms can be used without failure of a component leading to total loss of braking capability (the approach used for WSP systems).

13.4.2 TRAIN ACCELERATION PERFORMANCE

A related area that has not been considered in any depth by the author is the acceleration performance of trains. It is possible that a revised approach to train design could enable significant improvements in this area, with direct impact on the most significant station stop headway scenarios by reducing the station run out time and thus the time taken to clear overlaps / safety margins. For example, distributing traction motors through a train offers "spectacular benefits in acceleration, adhesion and redundancy" when compared with concentrating power in one or two power cars (Green 2000, p52).

Whilst the benefits to be gained by increased acceleration performance would not be as significant as equivalent increases in braking performance, the impact on capacity could still be significant for congested railways. This, therefore, represents another potential area for future research.

13.4.3 TRAIN BRAKING ALGORITHMS

During a lecture on the 'ethics and economics of speed signalling' in 1931, it was observed that "the braking problem is inseparable from the signalling problem and a mathematical conception of it is, or should be, one of the first steps towards a comprehensive solution of the signalling problem" (Crook 1931, p124-5). This observation is as relevant to the implementation of Automatic Train Protection systems as it was to conventional signalling based train control. With ATP, "if the safe protection distance is too long, it will reduce train operation efficiency and disturb normal train operation controlled by the drivers. If it is too short, train operation can not be ensured to be safe" (Bin 1996, p111). There is, therefore, a need to consider the assumptions relating to train braking performance made during the development of appropriate train braking algorithms.

The issues to be considered within this topic divide into two groups:

1. The general approach / philosophy to be adopted;
2. The development of specific algorithms for implementation.

Issues to be considered within the general approach include the impact of warning and intervention strategies:

- Whether to operate the ATP system with both service and emergency brake interventions, or to limit it to emergency brake intervention only;
- The types of constraint that should be considered within braking algorithms;
- The accuracy of trainborne odometry systems that will be used in determining speed and distance variables to be used within the braking algorithms;
- The need for contingency or safety margins within the algorithms;

These areas were touched upon during the author's research, specifically during consideration of speed supervision in section 6.3. Whilst the development of specific train braking algorithms is an implementation issue for any future ATP systems, to be based on the specific separation strategies and performance requirements of that system, further research into braking algorithms and means of enhancing the accuracy of trainborne odometry would be of great benefit when those systems come to be developed.

13.4.4 OPERATIONAL SUPPORT TOOLS

The discussion of capacity in section 2.5 showed that recovery margin and service control functions have a significant effect on a railway's achievable line capacity. During the literature review documented within this report, a number of significant opportunities were identified for the development of operational support tools that would enable the reduction of recovery margins and improved service control. These included:

- Automatic Train Operation for British main line railways
    - The development of image recognition systems;
    - Further safety analysis;
    - Identification of fields of application that would provide acceptable cost-benefit relationships.

  (Eberhardt 2001, p8; Postaire 1986, pp303-8)

- Automatic Train Supervision systems, including
    - Train path management systems;
    - Automatic implementation of train regulation strategies;
    - Decision support tools to advise operators on response to divergence from timetable, including resource control;
    - Automatic implementation of service control functions (amendments to timetable following disruption).

  (Winder 1999, p4; Hurley 1999, pp136-9 ; Sacha et al, pp 155-9)

It has been stated "If we do not introduce fully automatic unmanned trains on the railways others will do so for us, even if they have to invent a new form of transport" (Maxwell 1975, p229). Whilst this statement is now rather old, the recent reappearance of schemes for personal rapid transit systems would suggest that its sentiments still apply. Not only would the introduction of ATO and ATR provide opportunities for capacity enhancement, they may well be necessary for the mode to remain competitive in some market segments, particularly that of urban transport.

Central to the function of ATS is the concept of train regulation – a set of activities that will become more complex and critical with the introduction of ATO and moving block based control systems. However, the author has found very little evidence of research into optimisation of train regulation.

These areas of investigation fell outside the direct scope of the author's research into ATP and were not considered in detail. However, they offer the potential for highly significant improvements in capacity utilisation by implementation of techniques that do not appear to have been widely researched to date. As a result, they probably offer the greatest potential for further research of any areas identified by the author.

13.4.5 DEVELOPMENTS IN TIMETABLING AND SCHEDULING PRACTICE

Another significant operational issue can be found in timetabling and scheduling practices:

- Where UK railways tend to allow for train running times that are longer than required by the train's theoretical performance, add recovery time in the timetable or leave slots unused in congested areas (for recovery), European railways tend to use generous station dwell times (Green 2000, p52). Whilst the author's modelling did not specifically look at this, its results do appear to suggest that a stationary train in a multiple platform station is far less critical to overall throughput than one travelling at low speed along the main line. This suggests that the European approach is better for optimising capacity – whilst at the same time supporting punctuality;
- Repeated reference has been made within this thesis to the UIC's recommendation that only 75% of theoretical capacity should be utilised in practice, the remainder acting as recovery margin. Whilst this approach would support significant improvements in

punctuality and reliability when compared to current UK timetabling practice, it is possible that higher utilisation could be achieved reliably in peak periods if adequate additional margin were added either side of the peak;

- Variation in performance characteristics between consecutive trains also wastes capacity. "Railtrack estimates that the congested Brighton main line could take an additional three trains an hour if all trains had common characteristics" (Green 2000, p52). Similarly, traffic through the Channel Tunnel includes a nominal 20tph, based on 26-minute duration Eurotunnel car shuttle paths. A Eurostar actually requires 2.66 of these paths for a 21-minute journey; two flighted Eurostars take 3.66 paths between them; one freight train takes 1.66 paths for a 28-minute journey, 2.33 paths for a 30-minute journey or 3 paths for a 32-minute journey (Goldson 2003, p8). Capacity utilisation can therefore be significantly enhanced by flighting services of similar performance in the timetable. If the Eurostar trains were slowed down further through the tunnel and freight speeds increased, the capacity could potentially be increased even further (at the expense of slower passenger journeys and the need to invest in faster freight trains).

The impact of operating trains with different characteristics has not been considered in this study, beyond consideration of different stopping patterns through stations. Similarly, the subjects of flighting and recovery margins have only been touched upon briefly. Further research into these and other aspects of timetabling and scheduling would offer the potential for more robust and safer railway services.

13.4.6 TARGETED MOVING BLOCK

Analysis of Targeted Moving Block in chapters 10 and 11 confirmed that localised application of moving block to single in-line platform station areas would obtain all of the capacity benefit to be gained from full moving block application. However, this left three questions that were not answered by the author's modelling:

- Whether the cost and complexity of localised moving block would be significantly different to those of a full moving block system;
- Whether any potential cost / complexity savings would be worth the loss of full moving block's inherently improved perturbation response and;
- The level of increased risk due to system novelty (when compared to the proven technology of full moving block implementation).

Whilst the answers to these questions fell outside the scope of the author's study, they would certainly warrant future investigation in order to determine whether efforts to develop targeted moving block systems would be worthwhile.

13.4.7 EFFICIENCY, SAFETY AND DRIVING QUEUES

In section 12.5.2, the potential for in-cab signalling systems to extensively assist the driving task was identified. In order to make best use of this potential, the author proposed that in-cab systems should:

- Be developed with the clear purpose of providing driving and regulation cues in support of efficient railway operation and;
- Avoid the temptation to integrate in-cab signalling functionality with that of the train protection system.

As this topic fell outside of the author's specified area of investigation, he was unable to include a detailed analysis within his research programme of the pros and cons of such an approach, or to develop the details of how such systems should ideally be implemented. If the best is to be achieved from drivers and regulation systems in the future, there is clearly a need to consider the author's suggested approach to in-cab signalling implementation further, along with any alternative ideas for ways to improve signal clarity, enhance regulation and provide clear driving cues whilst avoiding problems of attention conflict and safety approval.

13.5 TRAIN SEPARATION STRATEGIES

13.5.1 TERMINAL STATIONS

During the author's analysis of elemental headway scenarios, the subject of terminal stations was not considered beyond the description and derivation of equations in chapter 8 and Appendix C (mainly because they represent a special case of the multiple platform station and diverging / converging junction scenarios). As most railway lines include a terminal station, the most direct continuation of the author's work would be to implement the terminal station equations within a model and determine whether this additional scenario impacts the general results obtained by the author.

This analysis could also be taken further in order to determine optimal point layouts and speed restriction profiles on approach to terminal stations or, indeed, whether such optimisation is required. Some research work and practical application have already been carried out on the use of a stepped speed approach to optimise terminal throughput, and this would be useful as a starting point for further research (Goodman et al 1999, p319; Nakamura 1999, pp324-7).

13.5.2 JUNCTION ARRANGEMENTS

The consideration of optimal layouts would also be of benefit in other junction areas (particularly where junctions are located in or near to station areas). During the author's research, typical layouts were assumed. However, it is likely that more optimal layouts could be found from the perspective of capacity, flexibility and reliability. The author is not aware of any significant research into this subject which, based on a review of some layouts implemented in recent years on UK main lines, would not appear to be well understood within the railway industry as a whole (as an example, see Cooksey 2001, p2-3)

13.5.3 PERTURBATION RECOVERY

The author has looked mainly at the impact of different train separation strategies and ATP system types on steady state railway capacity. However, in practice, a significant element of any railway system's achieved capacity will be determined by its response to, and recovery from, perturbations to the planned service.

The author noted in chapter 10 that moving block systems would be expected to achieve improved perturbation recovery when compared to fixed block. Similarly, an intermittent balise-based system would be expected to suffer additional delays when compared with both a conventional lineside signalling approach and a continuous transmission system, whether radio or localised loop based, when the service becomes perturbed. The modelling of perturbation recovery in order to verify (or refute) the author's expectations and assess the extent of any difference between ATP system types would be a significant area for further research following on from this thesis. In addition to the scenarios considered by the author within this thesis, the impact of areas with bi-directional working may also be significant for the study of perturbation recovery and would also warrant investigation.

To date, the author is only aware of one study conducted into perturbation response of moving and fixed block systems. This was a PhD thesis by Mr Hiroto at Birmingham University, submitted in 1997. Mr Hiroto limited his study to plain line sections and to comparison of fixed block with continuous in-cab ATP. He also neglected consideration of system delays. However, his thesis would provide an excellent starting point for further investigation of this subject.

13.5.4 MODELING OF SPECIFIC RAILWAYS

The research conducted by the author during preparation of this thesis was deliberately kept generic. However, further work to model specific examples of actual railways would now be beneficial in order to determine:

- Whether the generic findings are consistent with the practical experience of railways with the ATP system that they currently use (if any) and;
- To assess the impact of each alternative ATP type on the railway's particular layouts and service patterns.

It was unfortunate that, at the time the author commenced his research, there was no commercially available railway simulation software available that could compare the different ATP types that were proposed for consideration and their inherent system delays. At the time of writing this is still the case. However, some proprietary simulation packages capable of such detailed analysis do now exist and it would, therefore, seem likely that the continuing development of commercially available software will also enable its use for such analysis in the near future. When this becomes the case, the author believes that the use of such tools would enable a far more comprehensive study of ATP system impacts of specific railway environments.

REFERENCES AND BIBLIOGRAPHY

Entries in *italics* have not been referenced within this thesis, but are included as a part of the wider bibliography.

Abuzeid M R, 'Advanced Braking Control Strategies for Trains', PhD Thesis, Dept. of Electronic and Electrical Engineering, Loughborough University, Loughborough, UK, 1996

Adeney W, 'Best Practice in Station Dwell Time Management', IIR Rail Capacity Conference, 19-20/02/01, IIR, 2001

Ainoussa A, 'A Crashworthy High Speed Aluminium Train: The West Coast Main Line Class 390 Tilting Train' What can we Realistically Expect from Crashworthiness? Improving train design to withstandfuture accidents, IMechE, 2001

Ainscough T, Hesketh G, 'Non-Wearing Braking (A Blend in the End?)', Brakes and Braking Systems Seminar, IMechE, 1993

Alcatel, 'Keeping the Future on Track', IRSE Younger Members Conference Technical Visit, 1996

Aldritch M, 'Combating the Collision Horror: The Interstate Commerce Commission and Automatic Train Control, 1900-1939', Technology and Culture, Volume 34 No. 1, pp49-77, The Society for the History of Technology, January 1993

Allan J, Williams J, 'The Fundamentals Of Systems Engineering In Major Railway Projects' IRSE, 1995

Allan J, 'Systems Engineering', IEE Power Division Fourth Vacation School On Electric Traction Systems, IEE, 1997

Allen A, 'Introduction to Transmission Based System: Issue 2', WSL, 1996

Allen J, 'Crashworthiness Progress on London Underground', Railtech '96, Railway Rolling Stock, IMechE, 1996

Alston L L, Birkby J W, 'Developments in Train Control on British Railways', IRSE, 1971

Alstom, 'Shopping Trip Triggers Big Invention', Signalling News, Alstom Signalling Ltd, August 1999

Anderson E, Bahr H V, Berg E, Bjorck H, 'Stainless Steel Carbodies for High Structural Integrity and Safety', Structural Integrity and Passenger Safety, Southampton Boston: WIT Press, pp31-47, 2000, ISBN 185312 7841

Anderson E, Bahr H V, Nilstam N G, 'Allowing Higher Speeds on Existing Tracks – Design Considerations of the X2000 Train for Swedish State Railways', Proceedings of the Institute of Mechanical Engineers, Part F, Journal of Rail and Rapid Transit, Vol. 209, No F2, IMechE, 1995

Andrews D, 'Junction Signalling: Memorandum PP/avw/RBI', Railtrack, 1995

Anon. 'Moving Block: Is It Worth the Money', Modern Railways, November 1967 pp606-609, 1967

Anon(1), 'Product Specs: AATC', Harmon Industries Inc, 1999

Anon(2), 'Bart's Advanced ATC Nears Completion', International Railway Journal and Rapid Transit Review, Vol XXXIX No.9, pp17-19, September 1999, 1999

Anon, 'Train Performance Team: 2610a', AEA Technology, 2000

Anon(1), 'Railtrack West Coast Route Modernisation Functional Specification: RT/WCRM/PIMS-PO/200 Issue 2.5', Railtrack, 2002

Anon(2), 'Professional Driving TOC Policies (info from TOC'c & ATOC at 19/4/2002)', Railtrack, 2002

Anon(3), 'SSI Application Manual, SSI Data Preparation, Timing Constraints on Interlocking Data Complexity: SSI 8003-52, Issue 6', Railtrack Plc, 2002

Anon(4), 'Advanced Automatic Train Control', Fedral Transit Administration, printed from http://www.fta.dot.gov/library/technology/attc.html 25/02/2002, 2002

Anon(5), 'Data Entry for HST ATP Supervised Service Running: Cat No 56/057047, Issue 3 rev C' First Greta Western, 21st January 2002

Appleby G, 'Structural Review of the Timetable', Railtrack, 2002

Arifin J, 'Management of Track Access Using a System Approach - Transfer report', Sheffield University Department of Mechanical Engineering, 2000

Bailey C (Editor), 'European Railway Signalling, Chapter 9: Automatic Train Protection and Control' (Pages 265 to 281), A&C Black, 1995

Bainbridge L, 'Ironies of Automation', Automatica 19, pp755-779, 1983

Bamford P, 'Coding And Error Detection', System Engineering For Train Control, Alstom Signalling Ltd / Orchard Consultancy Ltd, 1999

Barber S, 'Capacity through Access to the Track' presentation slides, AGRRI Network Capacity Seminar, 2003

Barr A, 'Meeting the Customer Perspective', First RIA Traction and Rolling Stock Course, 1999

Barnard REB, 'SELCAB Automatic Train Protection for British Rail's Chiltern Lines', Aspect '91 International Conference proceedings, IRSE, 1991

Barnard REB, Uebel H, 'SELCAB Automatic Train Protection System', IRSE, 1992

Barnard R E B, 'Presidential Address - Glimpses into the Future', IRSE, 2001

Barnard(1) R E B, 'Train Protection Principles: Tilt Authorisation and Speed Supervision (TASS): C80056/TECH/00/00011/02', Alstom Signalling Ltd, 2002

Barnard(2) R E B, 'System Specification: Tilt Authorisation and Speed Supervision (TASS): C80056/SYSREQ/00/00001/05', Alstom Signalling Ltd, 2002

Barnard R E B, 'ERTMS-Compatible Control of Tilting Trains', IRSE News Issue 85 June 2003, IRSE, 2003

Barter WAM, 'Train Planning in the Railway Business Process', IRSE Convention, Lisbon, 2001

Barton D, Hollands R D, 'Mass Transit Railway Automatic Control Systems', Railway Signalling & Telecommunications Course, RIA, 1987

Barwell F T, 'Automation and Control in Transport', Pergamon Press, ISBN: 0-08-026712-2, 1983

Basu S, 'Tailoring a Radio Based Automatic Train Control (ATC) / Automatic Train Protection (ATP) System With Existing Train Traffic Control System and the Brake and Traction Control System of the Rolling Stock', Seminar on Modern Signalling and Telecom Systems for Developing Railways, Institution of Railway Signal and Telecommunication Engineers (India), 1998

BBC, 'Centrifugal Force', www.bbc.org.uk/dna/h2g2/A597152, printed 15th June 2004

Beilby D, 'Optimising The Benefits of Electric Braking', Brakes and Braking Systems Seminar, IMechE, 1993

Beilby D, 'TCS Characteristics Survey', Bombardier, 2001

Bergmann D R, 'Generalised Expressions for the minimum time interval between consecutive arrivals at an idealised railway station', Transp. Res., 1972, 6, pp327-341

Berndt P, Kroger U, Saumwebber E, 'The Principles of Operation of Track Brakes (Magnetic and Eddy-Current) – Interaction With the Track', International Conference on Railway Braking, IMechE, 1979

Bin N, 'Analysis of train braking accuracy and safe protection distance in automatic train protection (ATP) systems', Computer aided design, manufacture and operation in the railway and other advanced mass transit systems, Computers in railways Vol 5/ V1, pp 111-118, 1996

Binard C, Van de Voorde W, 'ACEC Transport's Automatic Train Protection System', IRSE, 1992

Blackall A T, 'Presidential Address',The Institution of Signal Engineers Proceedings: Inaugural Session, 1910

Blakeney A, 'Resistance of Railway Vehicles to Derailment and Roll-Over: GM/RT 2141 issue two', Railway Group Standard, Railtrack, 2000

Blakeney(1) A, 'Calculation of Enhanced Permissible Speeds for Tilting Trains: GC/RC 5521 issue one', Railway Group Standard, Railtrack, 2001

Blakeney(2) A, 'Recommendations for systems for the supervision of enhanced permissible speeds and tilt enable: GE/RC 8517 issue one', Railway Safety Approved Code of Practice, Railway Safety, 2001

Blakeney A, 'Automatic Train Protection: GE/RT 8039 draft 1j', Railway Group Standard, Railway Safety, 2002

Boocock D, 'Permissible Track Forces for Railway Vehicles: GM/TT 0088 issue one', British Railways Board, 1993

Bott K, 'TCS Operational Specification: TCS/JPO/RES/99/00089/02', WCML TCS Joint Project Team, 2000

Bott K, 'Design Rationale for TCS Operational Specification, TCS/JPO/TRP/00/00169/01C' TCS Joint Project Team, 2001

Bound A F, 'A review of the Art of Signalling', IRSE Proceedings 1913 - 15, 1915 session pp14 - 48, IRSE, 1915

Boycott P, Berieau M, 'Assessing Cost Efficiency and Safety of the ERTMS Technology', 6th Annual Rail Safety Congress, AIC Worldwide, 1999

Brauer D, 'Train Operation in Relative Braking Distance', Future Trends in Signalling and Train Control, IRSE, 2001

Braun A, 'Braking of High Speed Passenger Trains', International Conference on Railway Braking, IMechE, 1979

Braun A, 'High Speed Braking Requirements and Service Experience', Brakes and Braking Systems Seminar, IMechE, 1993

BRB, 'Junction Signalling: Standard Signalling Principle No. 6', British Railways Board, 1992

Brearley S A, 'Railway Safety Through Passenger Needs', International Conference on Railways for Tomorrows Passengers, IMechE, 1993

British Rail, 'Brake Systems: Basic Principles, TSU/90/395', British Rail, undated

Broadbent H R, 'An Introduction to Railway Braking', Chapman And Hall Ltd, 1969

Brotzman M, 'Signalling in the USA' Railway Technical Web Pages, 20 June 2000

BSI, 'BS IEC 60050-821 International Electrotechnical Vocabulary – part 821: Signalling and Security Apparatus for Railways', British Standards Institution, 1998

Burrage K, 'Risk Management in Safety Critical Areas', IEE Power Division Sixth Vacation School On Railway Signalling & Control Systems, IEE, 1996

Cable B, 'Vacuum And Air Brakes', Eleventh RIA Motive Power Course, 1994

Calver W J B, 'The Great Western Railway Automatic Train Control' http://info.ex.ac.uk/wjbcalver/gwratc.htm, printed 18/9/01

Campion G, Vanbreusegem V, Bastin G, 'Modelling, Simulation and Control of the Traffic at the Terminus Stations of Urban Underground Railway Lines', IFAC Control in Transportation Systems, 1986

Carney R S, 'Track Alignment for Tilting Trains: A Case Study on the Constraints to Line Speed Enhancement', MSc (Eng) dissertation, Sheffield University, 2001

Catalis Rail Training, 'Intermediate Signalling Technology – Layouts: Aspect Sequence Charts - Examples', Catalis Rail Training, 1998

Catalis Rail Training(1), 'Intermediate Signalling Technology – Layouts: Aspect Sequence Charts', Catalis Rail Training, 1999

Catalis Rail Training(2), 'Intermediate Signalling Technology – Layouts, Signalling the Layout', Catalis Rail Training, 1999

Central Trains, 'Defensive Driving: CO/DS3104 Issue Two', Central Trains, 2000

Challis W H, 'The Signalling System of the Victoria Line: Instruction Manual', London Transport, 1976

Chandler M, 'Increasing Capacity on London Underground's Central Line', Railtex International Railway Engineering Conference, Railway Infrastructure, IMechE, 2001

Cheeswright P R, 'Overturning of Tilting Trains in Curves – Investigation and Recommendations for Control: RR/SAM/99/038 Issue 1.1', AEA Technology Rail, 1999

Chew T C, Troger L, 'Signalling and Train Control for Singapore North-East Line', IRSE, 2001

Clark G, 'Automatic Train Protection Systems In Use', IRSE Younger Members Second Railway Equipment Seminar: Train Protection Systems, IRSE, 1999

Clark G, Threlfall P, 'The Jubilee Line Extension', IRSE, 2000

Clementson B (Chairman), 'Lessons Learned From the Automatic Train Protection (ATP) Pilot Schemes in the UK', The Train Protection Steering Group, 2001

Clementson J, 'Derailment', MSc in Rail Systems Engineering, Module MPE6630, Sheffield University, 2000

Cobb S, Nichols S, Herr S 'Human Factors of Real-Time Information Systems: VIRART Report No. VIRART/96/135', British Rail Research, 1996

Codd H A, 'Geographical Circuit Technique', IRSE, 1958

Coleman V, 'Railway Safety – Taking Stock', IEE Railway Group, 1999

Component Part, 'Braking', Modern Railways, February 2002

Connor P R, 'Single Line Operation', Railway Technical Web Pages, 9 November 1999

Cook R, 'System Architectural Implications Following SIL Assignment – A Railway Signalling Example', SILs – Does Reality Meet Theory, IEE Seminar, 2002

Cooksey A, 'Implications for Signalling of the Ladbroke Grove Inquiry', IRSE, 12th December 2001

Cooper G, 'Capacity through Reliability' presentation slides, AGRRI Network Capacity Seminar, 2003

Cope G H (editor), 'British Railway Track: Design, Construction and Maintenance, 6th Edition', The Permanent Way Institution, 1993

Corrie J, 'The Principles of Train Detection', IEE Power Division Sixth Vacation School On Railway Signalling & Control Systems, IEE, 1996

Corser P D, 'Sunderland Direct, NEXUS Trainstop Approach Speed (reduced overlap) Overlap, Overrun and Acceleration Data: E80058/SPE/PRJ/0054', 2001

Cotton J H, 'Mail Rail: The Post Office Underground Railway Centralised Train Control System', IRSE Proceedings 1993/94, pp109-112, IRSE, 1994

Crawshaw G, 'System Aspects of Electric Braking', Brakes and Braking Systems Seminar, IMechE, 1993

Crook G H, 'Ethics and Economics of Speed Signalling', IRSE 1931 proceedings, pp122 to 167, 1931

Cox D R, 'Queues', Methuen & Co. Ltd, 1961

Cox S, Tait R, 'Safety, Reliability & Risk Management: An Integrated Approach', Butterworth-Heinemann, 1998

Curley J, 'Railtrack's Role In Improving Performance', Modern Railways, P48, May 2000, 2000

Cusk T, 'Headway Analysis for Tomorrow's Railway', IRSE Younger Members Section, 2001

Dapre S, 'Introduction to Signalling – Automatic Train Control ' Institution of Railway Signal Engineers, 1999

Dasi-Sutton A, 'Davies Endorses TPWS', Modern Railways, P10, April 2000, 2000

Davies D, 'Automatic Train Protection for the Railway Network in Britain – A Study', The Royal Academy of Engineering, 2000

Day K M, 'Future Railway Operational Control', British Rail Research, 1990

Dell R, 'Automatic Junction Working and Route Setting by Programme', IRSE, 1958

Demetrescu M, Pedroso G, Merlo A C, 'Communication Based Signalling System Moving Block Application at the Sao Paulo Metro', Aspect '95 International Conference proceedings, IRSE, 1995

De Vilder F, 'International Technical Committee Report No 4, Implications Of Applying Transmission Based Signalling', IRSE, 1998

Doherty A, 'System Engineering On The Central Line And Northern Line Projects - A Practical Approach', IRSE, 1995

Dorn M R, 'Jerk, Acceleration, and the Safety of Passengers', International Congress Railtech'98, Technology for Business Needs, IMechE, 1998

Dow G, 'Great Central', Ian Allan Locomotive Publishing Ltd, 1965

Doyle J B, 'Crash Design of Steel Bodyshells for Visgin', What Can we Realistically Expect from Crashworthiness? Improving train design to withstand future accidents, IMechE, 2001

Duckitt H, 'Coded Track Circuits', 1967

Dyson A M, 'The Development of Standards for Tilting Trains', Third RIA Railway Infrastructure Engineering and Management Course, 2000

Eberhardt M, 'Driverless Operation of main lines', Future Trends in Signalling and Train Control, IRSE, 2001

Edwards T J, Anderson M W J, 'Development of the Heathrow Express Crashworthy Vehicle', Railtech'96, Railway Rolling Stock, IMechE, 1996

ERTMS Programme Team, 'ERTMS: Towards a Better, Safer Rail System', Railway Safety and The Strategic Rail Authority, April 2002

ERTMS Programme Team, '2003/2004 Progress Report', Strategic Rail Authority, 2004

ERTMS Users Group(1), 'General Glossary' reference EEIG:96e399 version 2, EEIG ERTMS Users Group, 1998

ERTMS Users Group(2), 'System Requirement Specification' reference EEIG:96e236 version 5A, EEIG ERTMS Users Group, 1998

ERTMS Users Group, 'Description of Brake Curve Calculation' reference EEIG 97E881 version 5E, EEIG ERTMS Users Group, 2001

ERTMS Users Group(1), 'Description of the brake curve calculation: Reference EEIG 97E881, version 6', EEIG ERTMS Users Group, 2002

ERTMS Users Group[2], 'Annex: Refference EEIG:OAE881 version 6', EEIG ERTMS Users Group, 2002

Etschmaier M M, 'Operational Planning and Control in Transportation Systems', IFAC Control in Transportation Systems, 1986

Evans A W, 'Estimating Train Accident Risk After Ladbroke Grove', Modern Railways, Dec 1999 pp885-887, 1999

Evans A W[1], 'Fatal Train Accidents Due to Signals Passed at Danger', Modern Railways, Jan 2000 pp 51-53, 2000

Evans A W[2], 'Fatal Train Accidents on Britains Main Line Railways', Journal of the Royal Statistical Society, 163 vol 1, 2000

Evans A W, 'Fatal Main-line Train Accidents', Modern Railways, March 2001 pp23-27, 2001

Evans R M, 'Checking the Speed of Trains: GO/RT3253 Issue One', Railway Group Standard, Railtrack PLC, 1995

EWS, 'The Professional Driving Policy', EWS, Undated (faximile sent 2001)

Fairbrother A L, Ripamonti P, 'A Modular Approach to Doubling Line Capacity', International Conference on Railways for Tomorrows Passengers, IMechE, 1993

Fargher[1] M, 'GM/RT2045, Braking Principles for Rail Vehicles', issue two, Railtrack, 2000

Fargher[2] M, 'GM/RT2046, Braking System Requirements and Performance for Trains Which Operate Above 125 mile/h', issue one, Railtrack, 2000

Fenner D, 'Application of Train Protection to British Mainlines', IEE Power Division Eighth Vacation School On Railway Signalling & Control Systems, IEE, 2000

Fenner D C[1], 'The Development of The Train Protection and Warning System', LUL Signal and Electrical Engineers Technical Society, 25/01/01

Fenner D [2], 'The Train Protection and Warning System', IRSE News, Issue 72 May 2001, IRSE, 2001

Fenner D [3], 'The development of The Train Protection & Warning system', Signal and Electrical Engineers Technical Society, 23 January 2001

Fenner D, 'Train Protection', IEE Review, pp29-33 Volume 48 Number 5, September 2002, IEE, 2002

Fenton J, 'Automatic Flight Control System Design Considerations', IEE control engineering Series 57, Flight Control Systems Practical Issues in Design and Implementation, pp170-189, IEE, 2000, ISBN 0 85296 766 7

Fielding C, Luckner R, 'Industrial Considerations for Flight Control', IEE control engineering Series 57, Flight Control Systems Practical Issues in Design and Implementation, p1-53, IEE, 2000, ISBN 0 85296 766 7

Fielding C, 'The Design of Fly-by-Wire Flight Control Systems', United Kingdom Control Council, 2001

Fleming E, 'Lineside Signal Spacing: GK/RT0034 Issue 4', Railway Group Standard, Railtrack, 2000

Ford[1] R, 'TPWS', IRSE News, Issue 73 July 2001, page 12 IRSE, 2001

Ford[2] R, 'Great Heck – The Physics of Crashworthiness', Modern Railways, April 2001

Ford R, 'Informed Sources: Public Immune To Safety', Modern Railways, February 2002

Francis J D, 'From The Signal Box Window', IRSE, 1994

Fry D, 'Smart Sanders', Adhesion: Money on the Line, IMechE, 2000

Fukuda M, Watanabe I, Yuji H, Takashige T, Tomii N, Ikeda H, 'An Integrated Train Control System Based on Digital ATP', RTRI, http://www.rtri.or.jp/infoce/wcrr97/c375/c375.html, printed 28/02/02, 2002

Fulford C R, 'Low Speed Performance of Wheel Slide Protection Systems Under Low Adhesion Conditions', MSc dissertation, Advanced Railway Research Centre, Sheffield University, 1997

Fulford C, Tunley J, 'An Integrated Apprach to Adhesion Management', International Conference on Railways as a System, pp83-91, IMechE, 2001

Galivel C, Pore J, 'SACEM –"Systeme d'Aide a la Conduite a l'Exploitation et a la Maintenance" Paris RER Line A ATP System', IRSE, 1991

Garbutt P, 'World Metro Systems', Capital Transport Publishing Ltd, Second edition, 1997, ISBN 1 85414 191 0

GEC Alsthom, 'Vehicle Functional Specification: TPL25/C/068 Issue A', GEC Alsthom Transportation Projects Ltd and Ferima Transporte, 1991 (faxed by Terry M on 13/06/2002)

Geduhn N, 'The ZUB 100 Intermittent Automatic Train Control (ATC) System', Aspect '95 International Conference proceedings, IRSE, 1995

Geyer T A W, Chapman C P, Morris M I, Chislett P R, 'Improving Safety at Platforms', Railway Engineering, Systems and Safety, Railtech'96, IMechE, 1996

Gibson P A, 'Composition friction materials for the replacement of cast iron railway blocks', Railtech 1996, Railway Traction and Braking, IMechE, 1996

Gill D C, 'Computer-assisted design of optimised signalling layouts for Rapid Transit Railways', PhD Thesis, Faculty of Engineering, Birmingham University, 1986

Gill D C, Goodman C J, 'Computer-based Optimisation Techniques For Mass Transit Railway Signalling Design', IEE Proceedings-B, Vol. 139, No.3, 1992

Glover J, 'Principles of London Underground Operations', Ian Allan Publishing, 2000

Goldson R, 'Passenger Operator's Perspective of Network Capacity' presentation slides, AGRRI Network Capacity Seminar, 2003

Gomide F A C, Amaral W C, Cury J E R, Gimeno L, Andrade M L, Mendes M J, 'Modelling, Optimization and Control of Subway Systems', IFAC Control in Transportation Systems, 1986

Goddard E, 'Supervision and Operation of Mass Transit Systems', IEE Power Division Sixth Vacation School On Railway Signalling & Control Systems, IEE, 1996

Goddard E O, 'Overview of Signalling and Train Control Systems', IEE Power Division Fourth Vacation School On Electric Traction Systems, IEE, 1997

Goddard E, 'Signalling and Train Control Systems 1', Railway Systems Short Course, University of Sheffield and Bechtel, 2002

Goddard J, 'A transport High', Metro, Wednesday August 20 2003, p15, 2003

Goldson R, 'Passenger Operator's Perspective of Network Capacity' presentation slides, AGRRI Network Capacity Seminar, 2003

Goodman C(1), 'ATO, ATP and ATC', System Engineering For Train Control, Alstom Signalling Ltd / Orchard Consultancy Ltd, 1999

Goodman C(2), 'Principles Of Moving Block Signalling', System Engineering For Train Control, Alstom Signalling Ltd / Orchard Consultancy Ltd, 1999

Goodman C J, Murata S, 'Customer Focused Traffic Control: A Passenger Oriented Traffic Evaluation And A Simulation Based On-Line Traffic Optimisation for Metros', ASPECT 99, IRSE, 1999

Gostling R J, 'Future Developments in Railway Braking', Railway Braking Seminar, IMechE, 1986

Gostling R, 'WSP Systems in Railway Braking', Brakes and Braking Systems Seminar, IMechE, 1993

Gostling R J, 'Higher Speeds on Conventional Lines', International Conference on Railways as a System, IMechE, 1999

Green C, 'Operations: The Second Chance', Modern Railways, July 2000 pp 51-57, 2000

Grenier M, 'Development of Brake Control Systems for Railway Vehicles', Brakes and Braking Systems Seminar, IMechE, 1993

Grossin J, 'Technology Integration in Advanced Commercial Aircraft Cockpits and Operational Systems', pp435 to 451 of Control and Dynamic Systems: Volume 52 Integrated Technology Methods and Applications in Aerospace Systems Design, Academic Press Inc, 1992, ISBN 0-12-012752-0

Guieu B, Basso G, Poitrasson-Riviere G, 'SACEM for Hong Kong MTRC', Aspect '95 International Conference proceedings, IRSE, 1995

Guilloux J, 'Speed Control System on the SNCF', Aspect '91 International Conference proceedings, IRSE, 1991

Gupta K K, 'Development of New Train Control Systems and Their Application In Developing Countries', Seminar on Modern Signalling and Telecom Systems for Developing Railways, Institution of Railway Signal and Telecommunication Engineers (India), 1998

Haimes Y Y, 'Risk Modeling, Assessment, and Management' John Wiley & Sons, 1998, ISBN 0-471-24005-2

Hall S, 'Danger Signals', Ian Allan Ltd, 1987

Hall S, 'The History and Development of Railway Signalling', IEE Power Division Sixth Vacation School On Railway Signalling & Control Systems, IEE, 1996

Hall(1) S, 'The History and Development of Railway Signalling in the British Isles: Broad Survey, Volume 1'Friends of the National Railway Museum, 2000

Hall[2] S, 'The History and Development of Railway Signalling', IEE Power Division Eighth Vacation School on Railway Signalling & Control Systems, IEE, 2000

Hansen I, 'Capacity Increase through Optimised Timetabling' presentation slides, AGRRI Network Capacity Seminar, 2003

Hansen I, 'A new tool for the analysis of train punctuality based in train detection data', AGRRI Network Capacity Seminar, 2003

Hardy B, 'Paris Metro Handbook', Capital Transport Publishing', 1988 ISBN 185414 104

Hardy D J, 'Can Railways Compete with Roads', Proceedings of the IMechE Part F, Journal of Rail and Rapid Transport, Vol 211 iss 1, pp1-10, 1997

Harmer J, Turner K, Riley C, 'EuroRadio and the RBC', IRSE, 2002-11-05

Harris N G, Godward E W, 'Planning Passenger Railways', Transport Publishing Company, 1992, ISBN 0 86317 174 5

Harris N G, 'How Much is a Life Really Worth', Rail, Issue 317 Nov 1997 pp26-28, 1997

Hazards Forum, 'Safety-related Systems: Guidance for Engineers', The Hazards Forum, 1995, ISBN 0 9525103 0 8

Heard B D, 'Automatic Train Protection and the Operational Railway' Railway Engineering, Systems and Safety, Railtech'96, IMechE, 1996

Hill J, 'Sustainable Growth For Europe's Railways', IEE Review, pp37-40 Volume 48 Number 5, September 2002, IEE, 2002

Hiroto T, 'Dynamic Response of Moving Block Signalling Systems', PhD Thesis, Faculty of Engineering, Birmingham University, 1997

HMRI, 'Railway Safety Statistics Bulletin 1998/99 – Railway Accidents 1998/99 ', Her Majesty's Railway Inspectorate, 1999

HMRI[1], 'Basic Description of Train Protection Systems', Her Majesty's Railway Inspectorate, http://www.hse.gov.uk/railway/paddrail/tps.htm, 01/11/00, 2000

HMRI[2], 'Railway Safety Statistics Bulletin 1997/98 – Railway Accidents 2997/98', Her Majesty's Railway Inspectorate, http://www.hse.gov.uk/railway/rsb9798.htm, 01/11/00, 2000

HMRI[3], 'Railway Safety Statistics Bulletin 1999/2000 – Railway Safety 1999/2000 ', Her Majesty's Railway Inspectorate, 2000

HMRI, 'Railway Safety Statistics Bulletin 2000/2001 – Railway Safety 2000/2001', Her Majesty's Railway Inspectorate, 2001

HMRI. 'Railway Safety Statistics Bulletin 2001/2002 – Railway Safety 2001/02', Her Majesty's Railway Inspectorate, 2002

Hofestadt H, 'Field Element Control from Vehicles', Future Trends in Signalling and Train Control, IRSE, 2001

Holgate D, Walters R J, 'British Railways Automatic Train Protection Specification', IRSE Proceedings 1992/93, pp128-142, IRSE, 1993

Holgate D J, 'Realising the Full Potential of Transmission Based Signalling', Practical Solutions for Implementing New Signalling and Communication Systems, IIR, 1998

Holtzer J, 'NS Reizigers needs ETCS and new lines to handle growth', Railway Gazette International, September 1999

Horsey C I, 'Railway System Performance Improvements on an Existing Metro', MSc Dissertation, Dept. of Mechanical Engineering, Sheffield University, Sheffield, UK, 2000

Horsely D, 'A Systems Engineer's View of Wheelslide', Eleventh RIA Motive Power Course, 1994

Howarth D A, 'Capacity Achievement on Thameslink 2000', International Conference on Railways as a System, IMechE, 1999

Howker A C, 'Have We Forgotten The Driver', IRSE, 1988

Howker[1] A C, 'World Wide Signalling Practices', IEE Power Division Sixth Vacation School On Railway Signalling & Control Systems, IEE, 1996

Howker[2] A C, 'Signalling and Communications in the 21st Century', IRSE Younger Members', 1996

Hunt GA, 'The Track as a Structure', First RIA Infrastructure Engineering & Management Course, 1996

Hunt M[1], 'TCS Overspeed Analysis', TCS Joint Project Team, TCS/JPO/TRP/01/00017/01, 2001

Hunt M[2], 'Further Analysis of TCS Speed Control', TCS Joint Project Team, TCS/JPO/TRP/01/00061/01A, 2001

Hurley J, 'The British Rail Automatic Route Setting System', Aspect '91 International Conference proceedings, pp 334-342, IRSE, 1991

Hurley J, 'Optimized Control of Railway Operations', International Conference on Railways as a System, IMechE, 1999

Illingworth R, 'GM/RT2100, Structural Requirements for Railway Vehicles', issue two, Railtrack, 1997

Informed Sources, 'Hatfield – 115 to 0 in 17 seconds', Modern Railways, March 2001 p18, 2001

Informed Sources, 'ERTMS in the UK – Sliding Back Relentlessly', Modern Railways, June 2003

Informed Sources, 'ERTMS: Can it be Made Cost Effective?', Modern Railways, Vol 61 No. 670, July 2004

'Insider', 'The Price We Pay When Accountants Value Life', Rail, Issue 316 Oct 1997 p39, 1997

IRSE 'IRSE Technical Committee Report No. 2: Operational Availability Of Railway Control Systems', IRSE, 1993

IRSE(1), 'Review of Signalling Philosophy', IRSE, 2000

IRSE(2), 'Railway Signalling: A Position Statement', IRSE, 2000

IRSE(1), 'Mainline Railway Signalling in the UK - A Review', IRSE Signalling Philosophy Review, IRSE, 2001

IRSE(2), 'International Technical Committee Contribution to the IRSE Review of UK Signalling Principles', IRSE Signalling Philosophy Review, IRSE, 2001

IRSE(3), 'Report by Working Group1 Institution of Railway Signal Engineers – Signalling Principles', IRSE Signalling Philosophy Review, IRSE, 2001

IRSE(4), 'IRSE Signalling Philosophy Review Working Group 2 – Human Factors', IRSE Signalling Philosophy Review, IRSE, 2001

Jackson A A, 'The Wordsworth Railway Dictionary', Wordsworth Editions Ltd, 1997

Jakubowski R J, 'Results Of Distributed And Leaky Feeder Systems Tests At 800 MHz In The Washington D.C. Metro System Tunnels', IEEE, 1994

Jarvis A, 'Portrait of the Liverpool Overhead Railway', Ian Allan Ltd, 1996, ISBN 0 7110 2468 5

Jeffrey D, 'ATP And Resignalling Of The Central Line System', IRSE Younger Members Second Railway Equipment Seminar: Train Protection Systems, IRSE, 1999

Jeffries A E H, 'Control & Communication for Unmanned Transit Systems, Aspect '91 International Conference proceedings, IRSE, 1991

James R, 'Closing the Loop – Retardation Control', Brakes and Braking Systems Seminar, IMechE, 1993

Kawaguchi K, Watanabe T, Yamanaka A, Kakinuma K, Tamaoki S, 'Development of New Brake System and Control Method for Future Railways', World Congress on Railway Research, 1997, http://www.rtri.or.jp/infoce/wcrr97/D6491/D649101.jpg, printed 28/02/2002

Kecklund L, Ingre M, Kecklund G, Soderstrom M, Akerstrdt T, Lindberg E, Jansson A, Olsson E, Sandblad B, Almqvist P, 'The TRAIN-project: Railway Safety and the Train Driver Information Environment and Work Situation – a Summary of the Main Results', Signalling Safety Conference, IQPC, 2001

Kerr R, Szelke E, 'Artificial Intelligence in Reactive Scheduling, Chapman & Hall, ISBN 0 412 72900 8, 1995 (IEE 519.7:621.9 KER)

Khessib M S, 'Improved Energy Exchange by Central Train Control', PhD Thesis, Faculty of Engineering, Birmingham University, 1989

Kichenside G, Williams A, 'Two Centuries of Railway Signalling', Oxford Publishing Co., 1998

Kimbler D L, 'IDEF0 Models', IE340 Lecture Notes #4, http://taylor.cers.clemson.edu/ie340/files/340-4.htm, printed 18/05/01

Kletz T, 'An Engineer's View of Human Error', IChemE, 2001, ISBN 0 85295 430 1

Klingler C D, 'IDEF0', http://www2.S/SWPI/STARS/SPD/section_5_2_1.html, printed 18/05/01

Knowles T, 'Automatic Train Control for Heavy Mineral Railways', IRSE Proceedings 1993/94, IRSE, 1994

Knowles R, Hayes J, Binns C 'Life Cycle Costing for European Rail Safety Technologies, Train Control and Protection Systems Conference, IIR Ltd, 1999

Konarzewska-Gubala E, Rymarczyk M, 'The use of Repeat Step Simulation for Controlling the Transportation Process', IFAC Control in Transportation Systems, 1986

Kubie J, 'One Safeguard too Many', Engineering Management Journal, IEE, 2001

Kuhn M, 'Automatic Route Setting Integrated In a Modern Traffic Management System', International conference on Developments in Mass Transit Systems, IEE, 1998

Kumagai N, Hasegawa I, 'The Role of Brake Technologies and Optimising Brake Systems for Higher Speeds on Shinkansen and Conventional Vehicles', International Rail Conference on Better Journey Time, IMechE, 1996

Lange H, 'The INDUSI Inductive Train Controller', Rail Engineering – The Way Ahead Vol. 1, IMechE, 1975

Lawrence L S, 'Concepts in A.T.O.', Railway Engineers Forum Meeting, 11 March 1982, IEE, 1982

Leach M E (Editor), 'Railway Control Systems', A & C Black, 1991

Leigh[(1)] M J, 'Elecronic Braking', Eleventh RIA Motive Power Course, 1994

Leigh[(2)] M J, 'Brake Blending', Proceedings of the Institute of Mechanical Engineers, Part F, Journal of Rail and Rapid Transit, Vol. 208, No F1, IMechE, 1994

Leveson N, 'Safeware System Safety and Computers', Addison-Wesley, 1995, ISBN 0-201-11972-2

Lewis J, 'Moving Block Proposal', Alstom Signalling Ltd, 2001

Lewis J H, 'Structural Crashworthiness – Possibilities and Practicalities', Proceedings of the Institution of Mechanical Engineers, Vol 216 No F2, pp 117-121, 2002

Lewis R B, 'Track Geometry Condition Monitoring and Assessment', Third RIA Railway Infrastructure Engineering and Management Course, 2000

Lewis R, Fisher D, Senthan S, Hunt M, Filipec M, Marathu H, 'System delay factors from 'Performance Gap Analysis for the GATC Level 2 Subsystem Design: TCS/JPO/TRP/02/00084/01B', TCS Joint Project Team, Alstom Transport Information Solutions, 27 June, 2002

Le Vegue O, 'L2MU Solution: TCS/EXT/TNT/02/00028/01A', Alstom Transport Information Solutions, 2002

Lockyear M J, Norgrove, K, 'Transmission Based Signalling Systems', IEE Power Division Sixth Vacation School On Railway Signalling & Control Systems, IEE, 1996

Lockyear M J, 'The Application of a Transmission Based Moving Block Automatic Train Control System on Docklands Light Railway', International conference on Developments in Mass Transit Systems, IEE, 1998

Loganathan A, 'Trainstops', LUL Examination Study Group, 1993

LICS (London International Correspondence School), 'ICS Reference Library Volume 22: Brakes and Brake Rigging', London International Correspondence School Ltd, 1914

Lowe E, 'Minimising the Human Factors Risk', Signalling Safety Conference, IQPC, 2001

Lu, 'Energy Absorption Requirements for Crashworthy Vehicles', Signalling and Communication Systems, AIC Conferences, 12th May 1997

Lu, 'Energy Absorption Requirements for Crashworthy Vehicles', Journal of Rail and Rapid Transit, V0l 216 No F1, IMechE, 2002

LUL, 'Technical Conciderations Involved in the Determination of Starting Signal and Speed Checked Overlaps for Manually Driven Trains: DW/YC/3/6/DES7, London Underground Ltd, Undated

LUL[(1)] Rolling Stock Engineering Department, 'Brake Standard, Part 1. Brake System Performance (Passenger Stock)', RSE/STD/006 – Part 1 Issue A, LUL, 1991

LUL[(2)] Rolling Stock Engineering Department, 'Brake Standard, Part 2. Brake System Control (Passenger Stock)', RSE/STD/006 – Part 2 Issue A, LUL, 1991

LUL[(3)] Rolling Stock Engineering Department, 'Brake Standard, Part 3. Brake System Components', RSE/STD/006 – Part 3 Issue A, LUL, 1991

LUL, 'Requirements For The Production of Headway Curves', LUL Signalling Principles Course, 1993

LUL, 'Automatic Signalling, Book 25', London Underground Limited, Signalling Principles Course Notes, 1994

Lundberg P, 'Eurobalise Transmission System', IRSE, 2002

Lyon D, 'Review of Factors Governing The Permissible Speed of Trains: Report No. ITLR/T9179/001', Issue A, Interfleet Technology Ltd, 2001

MacDowell B, 'Train Brakes FAQ', http://www.spikesys.com/Trains/brk_faq.html, printed on 11/04/01, 1994

Maltby S, 'Optimisation Case Study', MSc Programme in Rail Systems Engineering, University of Sheffield, 1998

Mansfield D, 'Northern Line: Managing the Wheel-Rail Interface', International Congress Railtech'98. Infrastructure Issues, IMechE, 1998

Marks M, 'Provision of Overlaps, Flank Protection and Trapping: GK/RT0064', Railway Group Standard, Railtrack Plc, 2000

Maxwell W W, 'Automation of Urban Railways', Rail Engineering – The Way Ahead Vol. 1, IMechE, 1975

McCormick H, McKenna P 'Signalling and train Control Systems – Requirements of the Business and Operating Company', IEE Power Division Sixth Vacation School On Railway Signalling & Control Systems, IEE, 1996

McCullie, J K, 'Defensive Driving Policy', Chiltern Railways, 2000

McKenna P, 'The Needs of the Business and the Operator', IEE Power Division Seventh Vacation School On Railway Signalling & Control Systems, IEE, 1998

McKeown D, 'Developing The Philosophy of Signalling', IRSE, 2001

Means R, 'Personal Rapid Transit (PRT)', http://www.electric_bikes.com/prt.htm, printed 18/04/2001, 2001

Mellitt B, 'Signals For The Future?', IRSE, 1996

Metcalf R, 'Controlling the Speed of Tilting Trains Through Curves: GE/RT 8012 issue one', Railway Group Standard, Railtrack, 1999

Miles T, 'Pendolino Progress', Modern Railways, August 2001

Millard D, 'Delivering a Metro Train Service-Managing the Performance of Key Components', International Conference on Railways as a System, IMechE, 1999

Milroy A P, 'Aspects of Automatic Train Control', PhD Thesis, Dept. of Electrical and Electronics Engineering, Loughborough University, Loughborough, UK, 1980

Miribadi A 'Fault Tolerant Train Navigation Systems Using a Multisensor Integration Approach', PhD Thesis, Departments of Automatic Control & System Engineering and Mechanical Engineering, Sheffield University, Sheffield, UK, 1999

Mitchel I, 'Examining Safety with Changing Technology', 6th Annual Rail Safety Congress, AIC Worldwide, 1999

Mitchell I H, 'Signalling Control Centres Today and Tomorrow', IRSE, 2003

Moens G, Stokes R, 'Channel Tunnel Rail Link Signalling and Communications', IRSE, 2003

Moir I, Seabridge A, 'Aircraft Systems: Mechanical, Electrical, and Avionics Subsystems Integration', Professional Engineering Publishing Ltd, 2001

Monselet O, 'Safety Aspects of Train Speed Supervision', MSc Dissertation, Department of Industrial Technology, University of Bradford, 2001

Montadert A, 'Measuring and Analysing the Efficiency of an Automatic Train Protection System', Proceedings of ESREL, SARS and SRA Europe Annual Conference, pp949 – 955, 2000

Montagne S, 'Permanent Way for High Speed', Rail Engineering – The Way Ahead Vol. 1, IMechE, 1975

Montigel M, 'Principles of Signalling 1' System Engineering For Train Control, Alstom Signalling Ltd / Orchard Consultancy Ltd, 1999

Montigel M, 'Principles of Signalling 2' System Engineering For Train Control, Alstom Signalling Ltd / Orchard Consultancy Ltd, 1999

Moray N, 'Perception Attention Automation and the Perception of Signals', Signalling Safety Conference, IQPC, 2001

Morris F, 'Potential use of Automation to assist with Railway crisis Management', Arthur D Little, undated (BR days)

Morris S, 'Composition Friction Materials For Railway Braking', Eleventh RIA Motive Power Course, 1994

Mott D, Schmid F, Watson C, 'A Goodness Function for Terminal Layouts of Urban Railways', Draft submitted to Network and Spatial Economics, Klewer, 2003

Mott D B, 'Heavy Metros', Railway Signalling & Telecommunications Course, RIA, 1993

Murata S, Goodman C J, 'Optimally Regulating Disturbed Metro Traffic With Passenger Inconvenience in Mind', International conference on Developments in Mass Transit Systems, IEE, 1998

Murray C L, 'The Application of Hydrodynamic Brakes to Railway Vehicles', International Conference on Railway Braking, IMechE, 1979

Muttram R I, 'A Train Protection Strategy for the UK', IRSE, 2001

Nagase K, 'A Study of Adhesion Between the Rails and Running Wheels on Main Lines: Results of Investigations by Slipping Adhesion Test Bogie', Proceedings of the Institute of Mechanical Engineers, Part F, Journal of Rail and Rapid Transit, Vol. 203, No F1, IMechE, 1989

Nakamura H, 'A Minimum Headway Running Pattern on a Moving Block System', Aspect '99 International Conference proceedings, IRSE, 1999

Nash C, 'Capacity through Demand Appraisal', AGRRI Network Capacity Seminar, 2003

National ERTMS Programme Team, '2003 / 2004 Progress Report', SRA, 2004

Nichols S, Cobb S, 'Driver Behaviour: VIRART Report No. VIRART/96/134', British Rail Research, 1996

Nishinaga E, Evans J A, Mayhew G L, 'Wireless Advanced Automatic Train Control', ASME / IEEE Joint Railroad Conference, IEEE, 1994

Nock O S, 'Railway Signalling', A&C Black (Publishers) Limited, 1980

Nock O S, 'The relationship Between Signalling and Brake Power in the Handling of Modern Traffic', IRSE, 1949

Noffsinger J, 'Alternative Approaches in North America to ERTMS Functionality', Signalling 2001Conference, IQPC, 2001

Norton D J, 'Safety by Redundancy', IRSE, 1979

NTSB, 'Railroad Accident Report: Collision of Washington Metropolitan Area Transit Authority Train T-111 with Standing Train at Shady Grove Passenger Station, Gaithersburg, Maryland, January 6, 1996', National Technical Information Service, 1996

Obara T, Kumagai N, Takiguchi T, 'Development Of Hybrid Rail Brake', Proceedings of the Institute of Mechanical Engineers, Part F, Journal of Rail and Rapid Transit, Vol. 209, No F2, IMechE, 1995

Oliver A, 'JLE Signals: What Went Wrong', New Civil Engineer, 21 May 1998, pp14 to 16, 1998

O'Neill M, 'The Application of ETCS to Railtrack Infrastructure', MSc (Eng.) dissertation (draft), Department of Mechanical Engineering, University of Sheffield, 1999

Ohyama T, 'Adhesion Characteristics of Wheel / Rail System and Its Control at High Speeds', 4th RTRI Symposium, Quarterly Review of RTRI, Vol. 33, No. 1, February 1992

Pachl J, 'Railway Operation and Control', Institut fur Eisenbahnwesen und Verkehrssicherung, 2000

Pearson L V 'Moving Block Railways Signalling', PhD Thesis, Department of Electronic and Electrical Engineering, Loughborough University, 1973

Peigne-Stage G, 'Special Tracks', http://iww.transport.alstom.com/GSC/DOCPOWERWE/RailBasi.nsf/26c3696b94e05dba81256659006cafd1/03cc6c97c75f0043c125689e004934ca?OpenDocument, *Alstom Transport Information Solutions, Printed 18/06/02*

Perry C, 'A Technical Vision For Britain's Railways', Second RIA Railway Infrastructure Engineering and Management Course, RIA, 1998

Pincock A J, 'Safe passage to Europe – Eurostar's in-cab signalling system', Journal of rail and rapid transit, Vol 212 issue 3,pp235-251, IMechE 1998

Pope R, 'The Institution of Railway Signal Engineers Booklet No. 27, Signalling the layout British Railways Practice', IRSE, 1975

Pore J, 'European Standards', IRSE 1996

Postaire J G, Stelmaszyk P, Bonnet P, Deparis J P, 'A Visual Surveillance System for Traffic Collision Avoidance Control', IFAC Control in Transportation Systems, 1986

Pritchard C, 'Brakes and Wheel/Rail Adhesion', IMechE, 1979

Prime H A, Sujitjorn S, Goodman C, Mellitt B, 'Energy Reduction by Dynamic Control', IFAC Control in Transportation Systems, 1986

Profillidis V A, Railway Engineering', Second Edition, Ashgate Publishing Ltd, 2000, ISBN 0 7546 1279 1

Railtrack, 'Engineering Safety Management Issue 3, Yellow Book 3, Volumes 1 and 2: Fundamentals and Guidance', Railtrack, 2000

Railtrack, 'Railway Group Standard: GO/RT3000 Master Rule Book', Railtrack , 1999

Railway Safety, 'Railway Group Standard GE/RT 8000 Rule Book', Railway Safety, 2004

Rasaiah W G, 'Watershed UK Accidents – Lessons Learnt', Proceedings of the Institution of Mechanical Engineers, Vol 216 No F2, pp 109-115, 2002

Rayers F G, 'Presidential Address', IRSE, 1989

Reddy N, 'TCS Design Team Current Thinking: TCS Balise', WCML TCS Joint Project Team, 10 September 2001, 2001

Research & Development Directorate, 'Proceedings of the workshop on the Application of Expert Systems to Transportation', Transport Canada, 1985 IEE Shelf Mark 519.713:629.1RES

Research & Development Directorate, 'Expert Systems: Their Application in the Canadian Transportation System', Transport Canada, 1986 IEE Shelf Mark 519.713:629.1RES

Richardson A, 'Issues Arising from the Introduction of Higher Speed Trains on the East Coast Main Line', MSc dissertation, Advanced Railway Research Centre, Sheffield University, 1999

Riley C, 'Transmission Based Signalling – The Answer to Economic Uncertainty and Design Practicality', Aspect '99 International Conference, IRSE, 1999

Ring H, 'Cost Benefit Analysis of Managing Safety: Practicalities of Using Assessment Tools to Measure the Cost and Risks of Safety', 6th Annual Rail Safety Congress, AIC Worldwide, 1999

RLE, 'Schematic Layout Alignment Section 2', Union Railways, 2002

Robson D, 'Networker Crashworthiness Development', First RIA Traction and Rolling Stock Course, 1999

Rodgers S, 'An Introduction To Spread Spectrum Communications', LUL , 1995

Rose J, Fisher A, 'Flexible Automatic Train Control', IRSE, 1989

Rosenbrock H, 'Machines with a purpose', Oxford University Press, 1990, ISBN 0-19-856346-9

Rowbotham A J R, 'Introduction to Signalling – Railway Signalling ' Institution of Railway Signal Engineers, 1999

Rowe G, 'Customer Requirements for Mass Transit Trains', Brakes and Braking Systems Seminar, IMechE, 1993

Rumble F J, Young N C, 'Radio Systems For Railway Networks', Railway Signalling & Telecommunications Course, RIA, 1993

Sacha K, Misiurewicz P, Traczyk W, 'On-Line Railway Traffic Planning and Control System', IFAC Control in Transportation Systems, 1986

Samet I, 'Crash Design of the Gardermoen's Train', What can we Realistically Expect from Crashworthiness? Improving train design to withstand future accidents, IMechE, 2001

Saumweber E, Kroger U H, 'Eddy Current Brakes', Railway Braking Seminar, IMechE, 1986

Sawyer D A, 'Railways: What can we Realistically Expect from Crashworthiness', What can we Realistically Expect from Crashworthiness? Improving train design to withstand future accidents, IMechE, 2001

Scherp J, 'Network Capacity – a view from the European Commission' presentation slides, AGRRI Network Capacity Seminar, 2003

Schmid F, 'Introduction to Systems Engineering For Transmission – based Train Control', System Engineering For Train Control, Alstom Signalling Ltd / Orchard Consultancy Ltd, 1999

Schmid (1) F, 'Why Failures Occur in the Wheel Rail System', Chairman's Introduction, Proceedings of the seminar on Why Failures Occur in the Wheel Rail System, Advanced Railway Research Centre, 2001

Schmid (2) F, 'Timetabling, Network Modelling and System Architectures (Additional Material), MSC Programme in Rail Systems Enginmeering, Signalling and Train Control Systems Module, Sheffield University, 2001 (Draft)

Schmid (1) F, Harris N G, 'Additional Material: Encyclopaedia of Railways – Section 2. Operating The Railway: How to Make a Profit. Economic, Technical and Organisational Principles Underpinning Railway Operations', Railway Systems Short Course, University of Sheffield and Bechtel, 2002

Schmid (2) F, Scott J, Watson C, 'Rolling Stock and Traction' Railway Systems Short Course, University of Sheffield and Bechtel (Extract from Kemps Engineering Handbook), 2002

Schmid F, 'A systems Overview of Railways' Railway Systems Short Course, University of Sheffield and Bechtel (Extract from Kemps Engineering Handbook), 2002

Schofield[1] K, 'Friction, Wear, Slip and Slide: Braking Systems for Railways', Fundamentals of Railway Traction Systems, MSc Programme in Rail Systems Engineering, Sheffield University, 2000

Schofield[2] K, 'Getting a Grip', Adhesion: Money on the Line, IMechE, 2000

Scholes A, Lewis J H, 'Development of Crashworthiness for railway vehicle structures', Proceedings of the Institution of Mechanical Engineers, Vol 207 No F1, pp 1-16, 1993

Scholes A, 'A Review of British Standards For Vehicle Structural Requirements' What can we Realistically Expect from Crashworthiness? Improving train design to withstand future accidents, IMechE, 2001

Scott J W, 'Permanent Way' Railway Systems Short Course, University of Sheffield and Bechtel (Extract from Kemps Engineering Handbook), 2002

Sergeant V, 'Factors Affecting Adhesion Utilization', Railtech'96, Railway Traction and Braking, IMechE, 1996

Shooter A, 'Get a Grip on Leaves: The Guide to Running a Safe and Reliable Railway During the Leaf Fall Season, Issue 1.0', Director of Engineering Performance, British Railways Board, 1993

Short R C, 'Philosophy and Specification Of Safety Systems', First Railway Signalling and Telecommunications Course, RIA, 1987

Short R C, 'Fundamentals of Signalling and Train Control Systems', IEE Power Division Sixth Vacation School On Railway Signalling & Control Systems, IEE, 1996

Short R C[1], 'The Legislative Framework For Railway Signalling And Control', Aspect '95 International Conference proceedings, IRSE, 1995

Short R C[2], 'Probability and the Railway Signal Engineer', Aspect '95 International Conference proceedings, IRSE, 1995

Short R, 'Human Factors and Railway Signalling', Signalling Safety Conference, IQPC, 2001

Skovdahl O, 'Railway Safety Management and Cost Benefit Analysis', 6th Annual Rail Safety Congress, AIC Worldwide, 1999

Smith A J, Addressing The Human Factor – Ensuring Technological Advances Are Not Compromised By Poor Human Performance', Train Control and Protection Conference, IIR, 1999

Smith C, 'Highways In The Sky: Flight Control', Yale-New Haven Teachers Institute, printed from www.yale.edu/ynhti/curriculum/units/1988/6/88.06.08.x.html on 15/01/2004

Smith G, 'Effective Capacity Utilisation – Crucial for Freight Growth' presentation slides, AGRRI Network Capacity Seminar, 2003

Smith V H, 'Victoria Line Signalling Principles', IRSE, 1966

Smith V H, 'Fully Automatic Controlled Trains', IRSE, 1982

Sone S, Ashiya M, 'An Innovative Traction System From The Viewpoint of Braking', International conference on Developments in Mass Transit Systems, IEE, 1998

Stanley P W, 'The Influence Of Human Factors On The Performance Of Railway Systems', IRSE, 1996

Starling M, 'Station Safety Cases – The Hong Kong Approach', Practicalities of Maximising and Demonstrating Rail Product Safety Conference, IIR, 2000

State of the Art, 'Braking Systems', Modern Railways, April 1994 p221-223, 1994

State of the Art, 'Bogies', Modern Railways, March 1996 p171-174, 1996

State of the Art, 'Train Control Management Centres: State of the Art', Modern Railways, 55(596) May 1998 p341-343, 1998

Steer J, 'AGRRI Network Capacity Seminar' presentation slides, AGRRI Network Capacity Seminar, 2003

Stretton M, 'TPWS – timescales getting tighter', Modern Railways, P417, 1999

Sullivan, T, 'TPWS – gearing up to implement faster', Modern Railways, P907, 1999

Sutcliffe D, 'Thameslink Division – Disruption Strategy', Thameslink Project, 1993

Sutcliffe D, 'Stations & Route Operations Group sub Working Party: Station Control Centre Strategy Paper', RLE, 2004

Sutton A, 'The Development of Rail Vehicle Crashworthiness', Proceedings of the Institution of Mechanical Engineers, Vol 216No F2, pp 97-108, 2002

Symons P, 'Australasian Signalling', IRSE, 2002

Systra, 'Avant Project 07: Railway Signalling and Safety Report (Translation) IGC SC03/07/018', Eurotunnel, 1991

Taskin T, Allan J, 'Overview of Signalling and Train Control Systems', IEE Power Division Third Vacation School On Electric Traction Systems, IEE, 1995

Taskin T, 'Central Line Service Regulation Requirements Specification – Issue 5', London Underground Limited Trains Delivery Group, Technical Report 107932 Document Number TDG-TR-177-A5, 1995

TCS JPT 'TCS Glossary of Terms and Abbreviations', Ref: TCS/JPO/DAB/99/00003 issue 02, WCML TCS Joint Project Team, 1999

TCS JPT, 'Headway Analysis Tool: HAT-SPT0100027~02H', TCS Joint Project Team, 2002

Technology Primer [(1)], 'Signalling technology: a Brief History', Modern Railways, 54(582) Mar 1997 pp163, 165-166, 1997

Technology Primer [(2)], 'Brakes: technology primer', Modern Railways, 54(591) Dec 1997 p780-783, 1997

Terry N J, '1995 Thorrowgood Study Tour – Signals, Who Needs Them in North America?',IRSE Younger Members / M&NW section, 1996

Thames Trains, 'Train Driving Policy', Thames Trains, 1999

The Economist, 'How not to run a railway', http://www.economist.com/printerFriendly.cfm?Story_ID=433992&CFID=1484055&CFTC, printed 27/03/01, written 23/11/2000, 2000

The Engineering Council, 'Guidelines on Risk Issues', The Engineering Council, 1993

Thomas J, Coenrad W J, 'The European Train Control System', IRSE 1996

Thompson D R, 'Line Speed Improvement Studies', Railtex International Railway Engineering Conference, Railway Infrastructure, pp61-67, IMechE, 2001

Thomas J M, 'What is ATP?', Railtex International Railway Engineering Conference, Railway Safety, pp111-119, IMechE, 2001

Threlfall P, Patel S, Chew T C, 'JLE Signalling System - A Systematic Management View', Railway Engineers Forum, 1995

Tilly J, 'Level Crossings on Rural Railways: Can the Railway Industry Continue to Subsidise Rural Settlements?', IRSE, 2000

Tinkham C C, 'Railway Signalling in the United States', IRSE Proceedings 1993/94, pp113-121, IRSE, 1994

Tirovic M, 'Development of a Wheel Mounted Disk Brake For A High Speed Train', Railtech'96, Railway Traction and Braking, IMechE, 1996

Tisi[(1)] J, 'Sanders', Wheels and Axles, Cost-effective Engineering, IMechE, 2000

Tisi[(2)] J, 'Sanders', Adhesion: Money on the Line, IMechE, 2000

Tomlinson G W, 'Formal Evaluation of the BRB Automatic Train Protection Pilot Schemes Report', ATP Evaluation Steering Group, The Train Protection Steering Group, 2001

Troup R, 'Invest in Signals to Increase Revenue – The Secaucus Transfer Story', IRSE News, Issue 83 March 2003, 2003

Tunley J, 'Managing Low Adhesion for Improved Operational Safety & Performance', First RIA Traction and Rolling Stock Course, 1999

Turner K, 'Terminal Tracks – Managing the Risk: GC/RT5033 Issue One', Railtrack Group Standard, Railtrack, 1999

Tyrell D C, 'US Rail Equipment Crashworthiness Standards', What can we Realistically Expect from Crashworthiness? Improving train design to withstand future accidents, IMechE, 2001

Tyrell D C, 'Rail Passenger Equipment Accidents and the Evolution of Crashworthiness Strategies', Proceedings of the Institution of Mechanical Engineers, Vol 216 No F2, pp 131-147, 2002

Uebel H, 'Presidential Address – Future Railway Control', IRSE, 2000

Uebel H, 'Durchsatz von Strecken und Stationen', Signal + Draht, (90)4, 1998 (in German)

Uebel H, 'Mainline ATP/ATC Intermittent and Continuous Systems', IEE Power Division Sixth Vacation School On Railway Signalling & Control Systems, IEE, 1996

Uebel H, 'Transmission Based Train Control Systems', Aspect '91 International Conference proceedings, IRSE, 1991

Uff J, Cullen W D, 'The Joint Inquiry Into Train Protection Systems', Health and Safety Commission, 2001

UNISIG, 'ERTMS/ETCS Class 1: System Requirements Specification Subset-026, Issue 2.0.0', UNISIG, 1999

UNISIG, 'ERTMS/ETCS Class 1: Performance Requirements for Interoperability, Subset-041, Issue 2.0.0', UNISIG, 2000

UNISIG, 'ERTMS/ETCS Class 1: System Requirements Specification Subset-026, Issue 2.2.0', UNISIG, 2001

Van Bloemendaal R M, 'Signalling and Train Control Systems', Aspect '91 International Conference proceedings, IRSE, 1991

Virgin Trains, 'Professional Driving Policy', Virgin Trains, Undated

Waboso D, 'Managing The Interface', IRSE, 1999

Waboso D, 'The ERTMS Programme Team Final Report Issued to HSC – April 2002', Railway Safety, 2002

Wagner W A, 'Techniques for on-board Automatic Aid and Advisory for Pilots of Comntrol-Impared Aircraft', pp321 to 358 of Control and Dynamic Systems: Volume 52 Integrated Technology Methods and Applications in Aerospace Systems Design, Academic Press Inc, 1992, ISBN 0-12-012752-0

Waller J, 'Control Concepts in Automatic Rapid Transit Systems', Rail Engineering – The Way Ahead Vol. 1, IMechE, 1975

Walrand J, 'An Introduction to Queueing Networks', Prentice Hall Inc, 1988, ISBN 0-13-474487-X

Wang J X, Roush M L, 'What Every Engineer Should Know About Risk Engineering and Management', Marcel Dekker Inc, 2000, ISBN 0-8247-9301-3 U.S.A.

Wang X, Ning B, Bing Z, Cheng Y, 'The Advanced Automatic Control System For Train Operation – Research on Moving Autoblock System', IFAC 12th Triennial World Congress, Automatic Control in Transportation Systems volume 3, pp 153-156, 1993

Watkins M(1), 'Mobile Communications And Eirene 1' , System Engineering For Train Control, Alstom Signalling Ltd / Orchard Consultancy Ltd, 1999

Watkins M(2), 'Mobile Communications And Eirene 2' , System Engineering For Train Control, Alstom Signalling Ltd / Orchard Consultancy Ltd, 1999

Watson(1) C, 'Materials Selection Criteria for Rail Vehicle Body Shell Construction (Rolling Stock II)', Railway Systems Short Course, University of Sheffield and Bechtel, 2002

Watson(2) C, 'Railway Traction and Braking Systems' Railway Systems Short Course, University of Sheffield and Bechtel, 2002

Watson J, 'An Overview of Disk and Tread Brakes', First RIA Traction and Rolling Stock Course, 1999

Weber O, 'High Speed Traffic Signalling', Rail Engineering – The Way Ahead Vol. 1, IMechE, 1975

Webster B, 'Safety System Ruled Out For Virgin's New 125mph trains', The Times, 20/04/2001

Webster B, 'Cost Blamed for Delay to Rail Safety Systsem', Times on Line May 19 2004, www.timesonline.co.uk, 2004

Wei S T P, Weng L K, Chong C T, 'The Singapore Rail Network For The 21st Century – A Driverless Mass Transit Network', Aspect '99 International Conference proceedings, pp 279-390, IRSE, 1999

Weightman C, 'Block Systems', IEE Power Division Eighth Vacation School On Railway Signalling & Control Systems, IEE, 2000

Whalley R H, 'Railway Braking Systems', IEE Power Division fourth Vacation School On Electric Traction Systems, IEE, 1997

White C, 'The Role Of The Signalling Engineer In The Modern Metro', IRSE Younger Members Conference, IRSE, 1998

White C, Millard D, 'Metro Signalling Operations', IRSE, 1998

Wickens A H, 'The dynamics of Rail Vehicles – From Stephenson to Carter', Proceedings of the IMechE Part F, Journal of Rail and Rapid Transport, Vol 212 iss 3 pp209-217, 1998

Williams D J S, 'Equipment Design: Ensuring New Technologies Ease The Problems Of The Man-Machine Interface For Optimum Reliability', Train Control and Protection Systems Conference, IIR Ltd, 1999

Williams J, Allan J, 'Systems Engineering For Mass Transit Railways', International Conference On Electric Railways In A United Europe, IEE, 1995

Wilson G D, 'ATCS – A Modular Train Control System', IRSE, 1988

Wilson G, 'Modern Train Control In North America', IRSE, 1995

Wiltshire P, 'Alterations to Permissible Speeds: GK/RT 0007 issue one', Railway Group Standard, Railtrack, 1995

Winder K, 'The Needs and Expectations of Train Operators', AEA Technology Ltd, 1999

Wing P F, 'Training: Current Developments and its Role in Loss Control', IRSE, 1993

Wolmar C, 'Knee-jerk Reactions Over Southall Crash are Misguided', Rail, Issue 315 Oct 1997 p21, 1997

Woodbridge A W, 'Some Notes on Acceleration, Speed and Retardation of Trains and their Relation to Signalling', IRSE, 1938

Woodland[1] D: 'ERTMS Level 1 General Description: TCS/JPO/TNT/00/00215/01', WCML TCS Joint Project Team, 2000

Woodland[2] D: 'ERTMS Level 2 General Description: TCS/JPO/TNT/00/00260/01', WCML TCS Joint Project Team, 2000

Woodland D, 'The Introduction of New Risks to an Operational Railway', IRSE Proceedings 2000/2001, 2001

Woolford[(1)] P, 'Provision of Lineside Signals: GK/RT0032 Issue 2', Railway Group Standard, Railway Safety, 2002

Woolford[(2)] P, 'Lineside Signals and Indicators: GK/RT0031 Issue 4', Railway Group Standard, Railway Safety, 2002

Wright N, Hamilton A, 'ATP – The Train Operator's Perspective', IRSE, 2002

Wright N, 'Application of On-board Train Protection and Lessons Learned', IEE 10th Residential Course on Railway Signalling & Control Systems, IEE, 2004

WSL, 'LUL Jubilee Line Extension Signalling System Overview -First Draft', WSL, 1995

Wyatt R S, 'Speed & Route Signalling', IRSE Proceedings 1981/82, IRSE, 1982

Xue X, 'Some Aspects of Structural Crashworthiness of Rail Vehicles: Part II of A Study of Statistics of British Rail Accidents and Some Considerations for Structural Crashworthiness of Rail Vehicles', 2001Presentation for the Transfer of Candidature From MPhil to PhD, Department of Mechanical Engineering, Sheffield University, 2001

Xue X, 'A Statistical Analysis of British Railway Accidents: Part 1 of a study on statistics of British Rail accidents and some considerations for structural crashworthiness of rail vehicles', Transfer report of MPhil to PhD, Department of Mechanical Engineering, Sheffield University, 2002

LIST OF ABBREVIATIONS

a - Train's acceleration rate [m/s^2]
ABS - Automatic Block Signalling
Amax - Maximum acceleration due to traction [m/s^2]
ARS - Automatic Route Setting
ASD - Automatic Sanding Device
ATO - Automatic Train Operation
ATP - Automatic Train Protection
ATR - Automatic Train Regulation
ATS - Automatic Train Supervision
B - Train's braking distance [m]
b - Train's braking rate [m/s^2]
bbu - Brake Build Up delay [s]. Assumed as a nominal value to represent the combined effect of the brake commencement delay and subsequent brake build up rate [s]
bbu_{fe} - Brake build up time of following train's emergency brake [s]
B_c - Time for the message that will be transmitted by a balise or loop to change once the conditions for it to do so have been satisfied [s]
Bd - Service braking distance [m]
Bd_c - Lead train braking distance from its last reported location assuming that it was travelling at the crashworthiness limit speed, V_c [m]
Bd_e - Emergency braking distance [m]
Bd_l - Lead train braking distance from its last reported speed at its last reported location [m]
Bj - Distance between berth track location and starter signal [m]
Bjc - Distance between berth location and departure junction clearance point [m]
Bl - Train berth location in platform [m]
BR - British Rail (UK state railway, succeeded in the main by Railtrack)
BR ATP - An ATP system compliant with the British Rail ATP Specification produced in the 1980s
BV - BanVerket (Sweden)
CBA - Cost Benefit Analysis
CFL - Chemins de Fer Luxembourgeois (Luxembourg)
CP - Junction Clearance Point
d - Signal separation [m]
DB AG - Deutsche Bahn AG (German railway public limited company)
DB - Deutsche Bahne DB (German state railway, succeeded by DB AG)
DMU - Diesel Multiple Unit
Ec - Collision Energy [J]
EEB - Enhanced emergency braking
EK_F - Kinetic energy of a train at its final location [J]
EK_S - Kinetic energy of a train at its start location [J]
EMU - Electric Multiple Unit
EP - Electro Pneumatic
EP_F - Potential energy of a train at its final location [J]
EP_S - Potential energy of a train at its start location [J]
EPS - Enhanced Permissible Speed
ERTMS - European Rail Traffic Management System
ETCS - European Train Control System
FS_{conv} - The point at which a converging train would attain the maximum permitted speed having traversed a junction
FSK - Frequency Shift Keying
g - Acceleration due to gravity ($9.81 m/s^2$)
GPS - Global Positioning System
GWR - Great Western Railway
HMRI - Her Majesty's Railway Inspectorate (part of Health and Safety Executive)
H_n - Headway distance for an 'n' aspect signalling system [m]
Ht_n - Headway time for an 'n' aspect signalling system [s]. Additional notation may be included to differentiate scenarios. For example, $Ht_{n(\text{lower speed restriction})}$ signifies the headway time for an 'n' aspect signalling system when entering a lower speed restriction.
$Ht_{(nH)}$ - Headway time for a 2 aspect signalling system that includes 'n' home signals [s]. Additional notation may be included to differentiate between the home signals where 'n' is greater than 1

$H_{tn(m)}$ - Headway time for an 'n' aspect signalling system operating at reduced speeds such that the train's driver can treat it as an 'm' aspect system [s]

i - The number of the signal preceding the outer home

ICE - Institution of Civil Engineers

IEE - Institution of Electrical Engineers

ih - The number of the inner home

IMechE - Institution of Mechanical Engineers

IRSE - Institution of Railway Signal Engineers

L - Train length [m]

LEU - Lineside Encoder Unit

LNE - London North Eastern Railway

LRV - Light Rail Vehicle (such as a tram or those used on the Docklands Light Railway)

LUL - London Underground Ltd

M - Vehicle mass [kg]

m_1, m_2 - The mass of the lead and following trains involved in a collision [kg]

N - North

NMBS - Nationale Maatschappij der Belgische Spoorwegen (Belgium)

NSB - Norges StatsBaner (Norway)

O - Overlap length [m]

ÖBB - Osterreichische BundesBahnen (Austria)

oh - The number of the outer home

PSR - Permanent Speed Restriction

P_t - Trainborne ATP system processing delay [s]

R - Curve radius [m]

RAMS - Reliability, Availability, Maintainability and Safety

RENFE - Red Nacional de los Ferrocarriles Espanoles (Spain)

RIA - Railway Industry Association

R_L - Length of Speed Restriction [m]

S - South

S - Sighting distance [m]

S_A - The distance between the intended approach release point and the last junction protection signal. [m]

$Sg_{(n)}$ - Signal spacing for block section 'n' [m]

Sj - Distance between last junction protection signal and the junction clearance point

SNCB - Société Nationale des Chemins de Fer Belges (Belgium)

SNCF - Société Nationale des Chemins de fer Français (France)

SPAD - Signal Passed At Danger

St - Sighting time [s]

t - Time [s]

TASS - Tilt Authorisation and Speed Supervision (a system for supervising the speed of tilting trains through areas of Enhanced Permissible Speeds)

Tc - Time for signal aspect to change once the conditions for it to do so have been physically satisfied [s]

TCS - Train Control System

Tdc - Time for a change in state of the train detection system to be registered once the conditions for it to do so have been physically satisfied [s]

T_{FSP} - Time taken by a train to travel over a defined distance between two defined locations in accordance with its speed profile [s]

Tip - Margin between indication of the need to brake on approach to the limit of a movement authority and the actual permitted speed being crossed if the train speed is not reduced [s]

Tlu - Train location update interval [s]

Tlu_e - Train location update error [m]. This represents the error that is introduced to a train's estimate of its own location, due to factors such as slip / slide, between absolute position location devices

Tlu_t - Train location update transmission delay [s]. This represents the time delay between an actual location measurement and receipt of a location message by the central processing system. Therefore, it includes the processing time required on-board the train in order to determine train location, based on the raw data of speed measurement and location update balise messages, and prepare a message for transmission, together with the transmission delay

Tp - Margin between crossing the permitted speed curve on approach to the limit of a movement authority and issuing of a warning if the train speed is not reduced [s]

Tph - Trains Per Hour

Tr - Time taken for the route to reset behind a train at a junction [s]

TSR - Temporary Speed Restriction

Tw - Margin between warning and service brake intervention on approach to the limit of a movement authority, if the train speed is not reduced [s]

UIC - Union Internationale des Chemins de Fer (International Union of Railways)

Ule - Update location error [m]. The distance between an optimal intermittent ATP update point and the preceding transmission point (whether a balise or loop)

U_t - Continuous ATP system transmission delay, including both time for the updated message to become available once the conditions for it to do so have been satisfied and the actual transmission time [s]

V - Vehicle speed [m/s] or designed line speed [m/s] or maximum permitted speed [m/s]

V - The relative difference in speed between two colliding trains [m/s]

V_{act} - Actual speed of operation [m/s]

V_c - The crashworthiness limit speed for a defined lead and following train combination (that is, the maximum speed of collision at which their combined crashworthiness design would be expected to protect passenger and crew survival space) [m/s]

V_f - Maximum permitted speeds for the following train[m/s]

V_{ftd} - Speed of a leading train that has left a platform when the following train is ready to depart the same platform to follow it [m/s]

V_l - Maximum permitted speeds for the lead train

V_R - Speed Restriction [m/s]

W - Work done by braking and tractive forces [J]

W - The difference between the distance over which a train must travel below line speed as a result of the speed restriction and the train following headway distance on plain line at line speed [m]

WSL - Westinghouse Signals Ltd

WSP - Wheel Slide Protection

X - Separation between consecutive two aspect signals [m]

x - Train Location [m]

∞ - Infinity

PAPERS ARISING FROM THIS RESEARCH WORK

Formal Contributions to Professional Institutions / Societies, Conferences and Journals:

'The Introduction of New Risks to an Operational Railway', IRSE Proceedings 2000 – 2001, pp106 to 115, presented on 3rd April 2001

'The Introduction of New Risks to an Operational Railway', Signal and Electrical Engineers' Technical Society, presented on 19th April 2001. Included in Annual Proceedings 2000-1

F Schmid, 'Train Control Research In Europe', IRSE, 2002 {10% contribution}

'Railways and Their Control Systems', IRSE Younger Member's Short Papers Competition, presented on 24 April 2002 – 1st Place

'The Limitations of Train Braking Performance: Adhesion, Acceleration and Jerk', paper and presentation, The Railway Technology Conference at Railtex 2002, ICE, presented on 26 November 2002

'The Capacity of Railways and Their Control Systems', Signal and Electrical Engineers' Technical Society, presented on 23 January 2003. Included in Annual Proceedings 2002-3

'The Context of Railways and Their Control Systems', IEE Railway Professional Network, March 2003

'The Limitations of Train Braking Performance', IEE Railway Professional Network, March 2003

'The History of Automation in Train Control', IEE History of Technology Event, presented on 12th July 2003

'Railway System Life Cycles', ASPECT 2003, IRSE, presented on 22 September 2003

'The Capacity of Railways and Their Control Systems', ASPECT 2003, IRSE, presented on 24 September 2003

'The Potential for Increasing Railway Capacity through Enhanced Train Braking Performance', Evolution of Modern Traction, University of Liverpool with the Locomotive & Carriage Institution, presented on 08 November 2003, Included in seminar proceedings.

'The Limitations of Train Braking and The Potential for Increasing Railway Capacity through Enhanced Train Braking Performance', Signal and Electrical Engineers' Technical Society, presented on 19 November 2003. Included in Annual Proceedings 2003-4

'Railway Control Philosophy', IRSE, to be presented 13 October 2004 and included in the 2004-5 annual proceedings

'Speed Restrictions and Speed Supervision Criteria', Evolution of Modern Traction, University of Liverpool with the Locomotive & Carriage Institution, to be presented on 06 November 2004 and included in seminar proceedings.

'The History of Automation in Train Control', Signal and Electrical Engineers' Technical Society, to be presented on 17 November 2004 and included in the 2004-5 annual proceedings

'Terminology Used in the Automation of Signalling and Train Control Systems', Journal of Railway Operations Research, accepted 26th August 2003, publication date not set

'The Capacity of Railways and Their Control Systems', IMechE Journal of Rail and Rapid Transit, submitted in June 2003, publication date no set

Sheffield University:

'The Capacity of Railways and Their Control Systems', Transfer Report for Sheffield University, presented on 19 December 2001

'The Capacity of Railways and Their Control Systems', Signalling and Train Control Systems module, MSc in Rail Systems Engineering, Sheffield University, presented on 21 January 2002. Repeated in updated form on 20 January 2003 and 19 January 2004

'The Capacity of Railways and Their Control Systems', Rail Short Course, Sheffield University and Bechtel Ltd, presented on 25 March 2002. Repeated in updated form on 24 September 2002, 18 March 2003 and 25th September 2003

'The Capacity of Railways and Their Control Systems', HMRI Course, Sheffield University and HMRI, presented on 23 November 2002. Repeated in updated form on 15 July and 16 December 2003

Informal Contributions to Employers Forums, Professional Institutions and Societies:

'Train Detection Equipment – A General Introduction', Croxley Green IRSE Exam Study Group, presented January 2001

'The History & Basic Functionality of ATP Systems', Croxley Green IRSE Exam Study Group, presented January 2001

'ATC, ATO & ATR', Croxley Green IRSE Exam Study Group, presented February 2001

'Moving and fixed block Signalling', Croxley Green IRSE Exam Study Group, presented February 2001

'Railway Systems', paper and presentation, Croxley Green IRSE Exam Study Group, presented February 2001

'ATP Systems', half day lecture, Alstom Training Day, presented on 22 September 2001

'Railway Control System Capacity', Croxley Green IRSE Exam Study Group, presented November 2001

'Research Proposals for Optimisation of Automatic Train Protection Systems', presentation and paper to Alstom UK Engineering Budgets Workshop, presented November 2001

'The Principles of ERTMS Levels 1 & 2', Alstom lunch time forum, presented on 5 February 2002

'The Railway System & System Railway: Understanding Capacity', Croxley Green IRSE Exam Study Group, presented February 2002

'Train Control System Overview', IRSE Technical Visit to Old Dalby ERTMS Test Track, Presented March 2002

'The Limitations of Train Braking', Alstom lunch time forum, presented on 26 March 2002, repeated 12 April 2002

'Speed Supervision Criteria', Alstom lunch time forum, presented on 24 September 2002

'Railway Systems', London IRSE Exam Study Group, presented 06 January 2004

'Train Control Philosophy', Bechtel Signalling, Systems and Operations Knowledge Groups, presented 15 July 2004

APPENDIX A - INTERVIEW TRANSCRIPTS

Introduction

The following transcripts were taken during interviews by the author (either in person, by telephone, or by e-mail). Each question provided in advance is listed in **bold** type, with the appropriate portion of the responses immediately following it. Supplementary questions raised during the interview are included within the transcript as *italic* type. Responses were provided on the understanding that they will not be used for any purpose other than the authors study / research at the University of Sheffield.

The transcripts have been listed within this appendix in alphabetical order by surname.

Bob Barnard BSc CEng FIEE FIRSE (Immediate Past President), (Principal Consultant, Alstom Transport Information Solutions).

Questions Raised by D. Woodland on 01/07/2002.

Qu 1. I have been reading your 'Train Protection Principles' document for the TASS system. Within the document you quote performance characteristics for the Class 390 and Class 221. Where did the data quoted come from?

We asked the train manufacturers. The Class 390 data came from Alstom Transport (Birmingham and Preston), the Class 221 data from Bombardier (with acceleration data originating from ALSTOM Transport, the traction package supplier to Bombardier).

Qu 2. Based on the performance data for the class 390 quoted in your TASS Train Protection Principles document and the ERTMS Users Group data, I make it that we need to allow:

- **2 second for driver reaction to a warning (at least);**
- **1 second for delay in traction cut off;**
- **4 second brake build up delay (I assumed this to be a linear build up);**
- **Maximum 3% down gradient;**
- **Maximum acceleration 0.44m/s2 (below 70km/h);**
- **Brake Rate 9%g;**

On that basis, I make the potential overspeed for a class 390 to reach close to 8km/h above the warning speed by the time the driver has responded, the brakes have built up, etc. if all of the worst case conditions apply at the same time. This means that a margin between warning and brake intervention of at least 8km/h is required if the driver is to be allowed

to act on the warning without invoking an intervention under worst case conditions at low speeds.

Unfortunately, the TASS criteria defined in the train protection principles document for the Class 390 is a warning margin of 4.5mph (7km/h) and an emergency brake intervention margin of 6.1mph (10km/h) above the supervised EPS. That is only 1.6mph (2.6km/h) margin between warning and intervention. We appear to be setting margins that will virtually guarantee an intervention if warning speed is reached whilst the train is still accelerating. Can you explain the criteria selected?

Traction cut-off is nowadays jerk limited to a very low value. Depending on whether this jerk limit is applied on intervention by safety systems, traction power cut time could be extended.

Air brake build up usually has a short "dead time" followed by an exponential build up of effort. The effect is not too dissimilar to a linear build up.

We assumed in setting the margins that drivers should never deliberately accelerate when travelling above line speed. As a result, the worst case margins required only need to account for acceleration due to gravity during any driver reaction delays and the brake delay and build up.

I have been advised by the TCS Operations team that drivers are trained to accelerate at full rate up to a speed restriction, only easing off 2 to 3mph below the ceiling speed. They also need to maintain speed at the permitted speed, which may result in them needing to apply traction whilst the train is close to the permitted speed. Due to the granularity in control available, this can result in some overspeeding with traction applied, all as a normal part of controlling the train. It, therefore, seems likely that if the driver makes an error, some traction power could still be applied as the train goes over the permitted speed (and even by the time a warning is given).

I did not realise that was how drivers were trained to perform. If that is so, I think that the margins are probably not sufficient as currently stated. I will reconsider them. Prior to final implementation, we also intend to conduct extensive testing at Old Dalby in order to determine the margins that would work in practice.

In that case, what do you consider to be the supervision criteria that must eventually be met by TASS?

It is driver error that leads to overspeed. I think that you need to allow for 'drift' in speed on steep inclines that causes the train to exceed the ceiling speed, as the driver has not really done much wrong there. However, if the driver accelerates hard towards the ceiling speed, and is still accelerating when he reaches the warning speed, I think that intervention would be fair. The margin between warning and intervention on ceiling speed supervision therefore needs to allow for acceleration due to gravity only.

The margin between permitted speed and warning may need to allow for some traction acceleration.

Braking to a target needs larger margins, since the driver may not be doing anything wrong in braking late on a day with good adhesion. If the margins allow for that, they should also be ample to allow for 'drift' due to gravity. Again, I suspect that you do not need to allow for any traction acceleration though. If a driver was approaching a braking curve, he would be unlikely to accelerate up towards the ceiling speed. Good driving practice would lead to braking early and easing off on reaching the new target speed, or (in the worst case) continuing at the same speed until the braking curve is met. If the driver is accelerating into a braking curve, I think he has done something wrong. So, you don't need to allow for traction acceleration at the start of braking curve supervision either.

The only other criteria I think you would need to consider is that the driver should never receive a warning or brake intervention whilst the train's speedometer is still saying that the train is below the permitted speed. Railtrack were very keen on that for TASS.

Supplementary Question by e-mail on 25/11/04:

Qu 3. Following the implementation of TASS, have there been any problems with the small margin between warning and brake intervention?

I have not heard of problems with drivers being unable to respond to TASS warnings, but I think the TASS team have been generally impressed with the accuracy with which a Pendolino driver (i.e. even on a relatively unfamiliar type of train) can brake from high speed, and hit absolutely the signed speed value at the board marking the start of the PSR. So, it may be that we have actually had very few accidental warnings.

Interview by e-mail with:

Ken Bott BSc (Hons), CEng, CPhys, MinstP, MIRSE, FPWI (Railway Safety Consultant, AEA Technology Rail).

Questions Raised by D. Woodland on 26/04/2001. Responses given by K Bott on 04/06/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

1981-1983 **Research Officer, Training Pool, BR R&DD.** Four six month assignments within R&DD, plus general railway experience.

1983-1985 **Research Officer, Physics Group, BR R&DD.** Lab and railway testing of inductive loop technology. Trackside warning philosophy study. Personal study of cardiac pacemakers and railway EMI.

1985-1986 **Engineering Assistant, BR DM&EE.** One of a two man investigation service provided by the Environment Section, covering all aspects of rail vehicle based noise. Specialisation in driving cab noise measurement, categorisation and remedial measures.

1986-1990 **Senior Scientific Officer, Physics Group, BR R&DD.** Technical work on ILWS project, including liaison with developer, production of technical and project documentation, detailed input to design, user requirement and system operation.

1991-1994 **Principal Scientific Officer, Physics Group, BR R&DD.** Trackside safety consultant to BR/Railtrack for LUL JLEP including assessment of options and formulation of policy. Analysis of trackside fatalities to identify causation. Site trials of prototype ILWS equipment and further technical development

1994-1995 **Principal Scientific Officer, Trackside Safety Team, BR Research.** Infrastructure safety consultant to Railtrack for North London Line works associated with the Jubilee Line Extension project, including design review, method statement review, audit for standards compliance and snagging on site.

1995-1998 **Safety Consultant, Trackside Safety Team, BRR/AEA Technology.** Independent Safety Assessor (railways) for TPWS Trials Project. Served on Human Factors Engineering team in WCML TCS project. Managed and conducted 'Human Factors and the Driving Function' project which produced a full featured driving simulator and ten detailed reports. Assisted CIRAS project.

1998-present **Railway Safety Consultant, Signalling & Operations, AEA Technology Rail.** Continuing TCS project work in the areas of human factors (train driver, signaller, shunter and trackside staff), operations, simulation, safety and design issues. Member of the Parliamentary Advisory Council for Transport Safety (PACTS) Rail Safety Working Party. Independent Safety Assessment of DART project, Bombardier safety case for new Virgin Cross Country trains and Thameslink 2000 train protection strategy.

Qu 2. The terms 'Signalling', 'train control' and 'railway control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system?

I'd not heard of 'Railway Control'. Train Control sounds different to Signalling but in practice I don't think it is. Perhaps someone wanted to avoid the implication that Signalling needs signals.

Qu.3 Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

I'd not heard of ATR, but it sounds like a generic term for what ARS does.
Eventually I found NTSB/RAR-96/04 concerning a collision on the Washington Metropolitan Area Transit Authority on 06/Jan/96. It defines an ATC consisting of three subsystems, namely ATP, ATS and ATO.

Qu 4. How would you define the capacity of a Railway system (and what would you consider to be appropriate units for your defined measure of capacity)?

No standard answer on that either, but it sounds like trains per hour, i.e. technical or operational headways. Actual capacity depends on what trains you run, and their characteristics. Often its the way their characteristics vary from each other that matters, but it could be between that and the signalling characteristics (headway again). Doug Holgate did a paper on Capacity (of a TBS) and defines it as "the number of trains passing a given point in a given time".

Qu 5. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

Whether it operates as a 'background' or 'foreground' system. i.e. does it sit in the background and only make its presence felt when providing protection in the event of a driver error, or does it force its way into the foreground by threatening intervention prior to driver error? This depends upon the algorithm used by the ATP to calculate braking curves, and the target stopping point that it is aiming for. If the ATP is set the same target as the human then I think conflict is inevitable. Even where the targets are different (as with TPWS practice if not design) then the risk is still present. The Driver should be free to drive the train, not monitor an ATP system in order to avoid intervention. TCS/JPO/TRP/99/00035/01F includes some discussion of this area.

Qu 6. I would welcome any views that you may have as to ways in which Automatic Train Control systems could best be optimised to create the greatest benefit for UK railways in both the short and long term?

Following on from 5, don't de-skill the Driver or distract him from his proper and rewarding task.

Qu 7. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

Usual stuff such as ease of implementation (resources required, design effort, system downtime during installation), cost, RAMS, etc. It has to provide a justifiable benefit without imposing an unacceptable right side failure burden.

Qu 8. The introduction of higher speed operation, higher capacities or changes to track layouts may all offer operational benefits, will influence the commercial viability of projects and will also affect safety. How do you think the trade off between safety and operational flexibility can best be managed?

Trust the Driver to do his job. Protect his errors with ATP.

Qu 9. What information do you think:

A. Is required by a driver in order to perform the tasks asked of him

B. Would be beneficial to a driver to assist him in the performance of his tasks

C. Would be detrimental to a drivers ability to perform the tasks required of him

Carte à la horse. The Driver's information requirements are a function of the task. The Driving task is critically affected by the signalling system, e.g. route and speed signalling produce very different tasks and therefore information requirements. I edited a report on a possible route signalling Driver Information System back in 1996 (written by HF folk at the University of Nottingham). Paul Le Vesconte has a copy.

Qu 10. How do you think that your answers to question 9 would change if the system included ATO?

It depends upon what role the Driver would have under an ATO system. Humans generally speaking make rather poor 'monitors' because by the time they are aware that something is amiss its too late to do anything about it. To intervene effectively and take manual control of a situation the human must already have situational awareness, and that is very difficult if action is rarely needed.

Tim Brockbank HND (Senior Engineer, Alstom Transport Information Solutions).

Questions Raised by D. Woodland on 20/09/2001.

Qu 1. Can you confirm the timing of target speed indications to the driver with BRATP?

The BRATP systems were designed to give the driver an indication of an upcoming target speed 3 seconds running time at line speed before a warning for under braking would need to be given. At line speed he would therefore get 3 seconds warning, but if the train was travelling at a lower speed, the warning time would be longer.

Qu 2. Can you confirm the conditions under which the train trip function becomes active on BRATP following operation of the train trip override button?

The train trip function is re-enabled when the signal that was at danger has been passed or after the train has travelled 100m or if a subsequent green signal is passed.

Qu 3. Can you confirm the supervision criteria used for the partial supervision mode in BRATP?

It only provides speed supervision at the maximum permitted vehicle speed.

Qu 4. Can you confirm the indications given to the driver for operation in full supervision mode by BRATP?

In full supervision mode a green LED is lit at the appropriate point around the speedometer to indicate the current speed limit. When a lower speed limit is approached, the new target speed is indicated by the appropriate green LED flashing. Steady yellow LEDs located at lower speeds around the speedo is then lit in place of a green one to indicate a release speed. In addition to this, a three character dot-matrix display is used to give an indication of the supervised signal stopping point:

- '===' for no supervised stopping point;
- '..0' for stop at the next signal but two;
- '.00' for stop at the next signal but one;
- '000' for stop at the next signal.

A warbling alarm then sounds and the LED indications are extinguished during a brake intervention or warning. The dot-matrix display also changes to flash the current supervised speed, target speed or supervised signal stopping point to indicate the cause of the brake intervention.
A 'blip' is sounded in the cab whenever a change occurs in the displays.

Qu 5. Are all trains on the GWML and Chiltern Lines fitted with the relevant BR ATP system?

Not on the GWML ATP scheme. Chiltern Railways have fitted all new rolling stock and operate a strict 'no ATP, no go' policy in the areas of line fitted with ATP. Being the only operating company to use that part of the rail network brings with it the advantage of being able to operate a truly comprehensive ATP system.

Qu 6. I have heard that the BR ATP schemes suffered from poor reliability. What were the main causes of this?

The 165 fleet and HSTs of GWML were fitted with end mounted tacho generators, relying on mechanical rotation. These included bearings which become worn by the forces that applied to the train axels. As a result of that they were very prone to failure – and degrade rapidly once the bearings start to wear. That was the main source of unreliability that I am aware of.
The Chiltern Lines 168 fleet were fitted with a 2 phase speed probe (with a grooved collar on the axle end and two proximity sensors detecting the passing of the grooves). It does not rely on mechanical contact to measure rotation and has proved much more reliable, with comparable accuracy.

Interview by e-mail with:

Steven Brown BEng, AMIEE (Vision Project Engineer, AEA Technology Rail).

Question Raised by D. Woodland on 11/04/2001. Responses given by S Brown on 11/05/2001.

How would you define the capacity of a Railway system (and what units would you give your definition)?

I would certainly agree that capacity can be defined in different ways to the standard tph figure. My best response would be that the definition depends very much on the question that you are trying to answer.
A figure of trains per hour is useful to the railway network operator and as a design requirement, since it allows the design (by standard and established techniques) of, for example, the track layout and signalling system for a new or upgraded line. As you are no doubt aware, Railtrack uses capacity in tph as the standard in its "rules of the plan" documents.
Tph is however rather dependent on the type(s) of trains you are talking about, so it is important to consider this alongside the headway or tph figure.
However, from the point of view of the train operator, capacity is better defined as freight tonnes or passengers per hour. A good example of this is the provision of extended freight loops; the tph figure remains the same but the capacity of the route is increased since it is possible to operate longer and heavier freight trains (provided more powerful locomotives are available of course) Another example is the enlarging of the loading gauge
to allow double-deck trains.
There is also the issue of whether the capacity is sustainable; the UIC standard is that a railway should not be operated in excess of 75% of available capacity (where capacity is in tph) (UK can be up to 95%, e.g. Victoria). It is therefore necessary to distinguish between short-term or maximum capacity, and long-term or sustainable capacity. The problem on the WCML is not so much the peak capacity, as the fact that the line is effectively being used 24 hours a day for 5 days of the week.
Having said this, I would still regard headway or tph as the best measure of capacity from the operator's point of view, and it is the measure I am using for my benchmarking studies. I have not really come across any alternative definitions but if I do I will let you know.

Interview by e-mail with:

Michel Carnot Ecole des Mines de Paris (Senior Engineer for Singapore North-East Line Signalling Contract, Alstom Transport Information Solutions).

Question Raised by D. Woodland on 05/03/2003. Responses given by M Carnot on 05/03/2003.

Qu 1. I was wondering if you knew the frequency with which trains transmit their location signal in the Alstom SACEM moving block system implemented in Singapore?

On Singapore, train updates its location to the trackside ATC every 624 ms.

Aditional Question Raised by D. Woodland on 25/03/2003. Response given by M Carnot on 05/03/2003.

Qu 2. I am interested to know what size safety margin is required for moving block operation, and whether it is a fixed value or speed related. Can you give me any insight into the Alstom SACEM view on what is required?

The size of the safety margin is 15 meters. It is a fixed value.

Jim Carpenter BSc(Hons), MIEE, C.Eng, EurIng (Principal Safety Consultant, CFG Admiral).

Questions Raised by D. Woodland on 16/05/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I have worked for a company called Admiral Management Services, now called CFG Admiral for the last 12 years as a consultant on system safety. Prior to that I worked for over 20 years on developing major systems, mainly nuclear reactor protection systems for submarines. Since joining Admiral in 1989 I have done a succession of safety consultancy jobs in a number of industries – mainly air traffic control, defence and railways, with a small amount in petro-chemical and other industries. My work in railways in particular started in the early nineties. I was the independent safety auditor for Queensland Railway on an automatic train protection system that was being developed for them by Westinghouse, for 4 or 5 years. I also worked for the best part of a year with Railtrack IT services as about the time of privatisation, and we established a methodology for doing safety assessments of railway database systems. We then carried out a sample safety analysis of many of these. After that I briefly worked as part of the Translink consortium on the TCS bid.

I have now been with the JPT for just over a year doing safety assurance work.

My only involvement with the railways has been from the safety perspective. I have no other railway background.

Qu 2. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

The main issue is deciding what the true purpose of the ATP system is. It sounds as though it should be obvious, but it isn't. The first question is, is the ATP distinct from Cab Signalling? There is no reason why they should be linked, but unfortunately there is a growing body of opinion that says that ATP and Cab Signalling are somehow intimately related. The Railtrack standards definition of ATP describes the properties of an ATP system, but then also describes the fact that it provides in-cab signalling. In my opinion those are utterly different functions.

So, first of all what is the scope of the ATP? Having decided that (particularly whether or not it includes in cab signalling as a function), you then have to decide how the driver is going to drive using it. To what degree will the ATP system influence the way the driver drives. Conceptually it should be possible to have an ATP that sits completely in the background, giving the driver no indication of its behaviour other than demanding a shut down or providing failure information.

The links to in cab signalling often come about due to deficiencies in the way the ATP is being done. People say you have to give in cab signalling because the ATP will always be very restrictive and you don't want the driver to get irritated by it, switch it off and do other things like that. That is fair enough, but it is only true because it is a rubbish ATP, not because it is conceptually needed.

If you are putting in cab signalling, is it there to cover deficiencies in the ATP, or for some other reason. If it is for another reason, then make sure that is clearly understood up front.

My concern is that ATP systems often dramatically change the way drivers drive – particularly if you give an indication of the permitted speed. I think that is the most dangerous thing that you can do. It provokes a head down driving style. I am convinced that a proportion of drivers will then behave completely differently, particularly in conditions of adverse visibility.

If it is a pure ATP (i.e. no in-cab signalling), you need to be clear why it is there, and what you expect of it. Most people expect ATP to be an overlay on the conventional system that provides additional safety. Very often people make decisions that means it does not add additional safety, but ends up providing THE safety. People then abandon thinking about the safety role of the driver. We have a view in this country at the moment that ATP has to be SIL 4 – why? There is no logic to that. It stems from the fact that in cab displays corrupt the driver by showing target speeds. If the ATP was a black box with the sole purpose of providing safety, and it was a good system, you would not have to ask a driver to rely on it. Instead of being THE system to provide safety, it could then be a secondary system – which I am sure was the original concept.

Everything hinges on getting the concept right. I think it would be entirely credible to produce a pure ATP sub-system that was only SIL 2 – if you have not influenced the driver so that he still has the normal means of control. Railways run at the moment with a pretty good safety record with no ATP. So, providing you don't change the drivers driving style, putting any ATP on should add an improvement to safety. You could argue that an ATP is worth having if it prevents 9 out of 10 accidents, but we don't currently have that kind of thinking because we have coloured it completely by saying that it has to have an in cab signalling system. That means it has to be SIL 4, which makes it desperately difficult to produce.

In the Queensland Railways system, we argued that it was a SIL 3 system, on the basis that it in no way detracted from the drivers legal and moral responsibility to drive the train correctly. We removed the requirement to show permitted speed, which removed the 'benefit' a driver may gain from head down driving. You just give the driver a warning before intervention, and log the warnings so that action can be taken against any driver who has a high log rate – and clearly a bad driving style. If you ensure that the driver still has all of the responsibility and information that he needs to drive the way that he did before, then you can argue that the ATP is a supplementary system. I think we could probably have argued that it was less than a SIL 3 system.

The concept is currently that ATP must always be on the safe side, with large margins for safety. If that is going to irritate the driver, why not adjust the margins of safety, so that it is not claimed to be 100% safe and under certain conditions it will not guarantee safety, but if it prevents 999 out of 1000 accidents that would have occurred, that is still a useful improvement – and probably still means that you won't get any accidents in the life time of the system. So it could well be that we have ended up being overly conservative and expecting too much from ATP systems.

That argument seems sound for plans to fit ATP to a railway that does not currently have ATP. If the railway had a SIL 4 ATP to current practice, do you think that we could come back at a later date and replace it with a new ATP system that effectively has a lower level of safety?

No. We are setting a massive precedent at the moment.

The case against the precedent is to show that the original logic was wrong, and that you did not have SIL 4 as you thought you did. For example, if we put a SIL 4 system on to WCML, we may in the future be in a position to say that although we thought we were installing a SIL 4 system, all we actually achieved was SIL 3 or SIL 2. You might then be able to argue that you only produced a SIL 2 system that was good enough, then you could argue that SIL 2 again would be good enough.

I think the industry has got itself into a terrible state. It has not thought these things through, but has muddled concept after concept.

The rail industry seems to have forgotten what is normally regarded as good practice in safety related industries. It has watered things down to such an extent that people have become confused. A good example is that you should have inherent technical defences against systematic faults in systems. It is not enough to depend upon high quality design procedures (although they are important). At the end of the day, the system that you have designed should have real technical defences in it against so-called common cause failures – software errors, systematic design or manufacturing errors, EMC susceptibility. Problems where simple duplication will not solve the problem. That is stated in the CENELEC standard 50129, but it is only one sentence hidden away in many that are mostly recommendations rather than telling you what you have to do. Safety related, and safety critical systems in particular, should have defences against faults built in. Not just random hardware failures, but the inevitable systematic faults in complex systems.

One of the ways of doing this is the concept of diversity – having things done by different sub-systems in different ways. Not just replication of identical units, but achieving something in several different ways. In my opinion the best way of doing this (which is recognised by many different industries – but not the

railway industry) is to put in diversity at the highest possible level. The best way to do that is to separate control and protection systems.
In some cases you cannot do that. In a fly-by-wire aircraft there is no difference between protection and control – because there is no safe state. Where you can identify a relatively safe state for the whole system to be in, which for a railway could be stopped, it does become possible to consider having a control system (with all of the cleverness in it), and a protection system (that is relatively simple) which just forces the thing being controlled into its safe state.
Because these systems will have different specifications, they will be truly different. They will be as diverse as you can make them. I think this would echo across to the railway industry very well. It is telling you that the control of the train should be completely separate from the protection of the train. The thing that controls the train then has the primary responsibility for safety – it has to control it as safely as it can. It has the first responsibility for safety. The safety system should then be the safety net. You don't rely on safety nets. You don't use them all the time – you use them as little as possible. You don't rely on fire extinguishers as the way of putting your gas cooker out. You rely on being able to turn it off at knob. You have then split the overall responsibility for safety between two systems. Neither of them have to be that difficult to produce to get real benefits. It is entirely possible to have 2 SIL 2 systems that will be good enough.

Qu 3. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

I think that the thing that would actually have the greatest effect are the political perspective, as expressed in the Cullen report – which has already made it clear that they are expecting ATP to be spread. Also the high-speed rail directive will I think rapidly extend to all high speed rail lines and is likely to spread to all other lines too.
Having some already approved kit will assist enormously the take up by railways.
What you haven't asked is what I think is discouraging wider / quicker application. One of the things at the moment is actually appreciating the fact that the ERTMS specification is both an enormous help, but also an immediate hindrance. clearly once an ERTMS specification is truly accepted and issued, that will be of great value, at the moment it is only a provisional specification in a state suitable for trials. It is highly likely that it is going to go through a substantial period of change. We have a specification at the moment that is not supported by any reasonable safety analysis or human factors analysis. It is riddled with errors and silly mistakes. So actually building real kit that uses it is a problem. People are not sure when this problem is going to be resolved. It is at least 2 years away before we get a more robust version.

Qu 4. As an additional point to this, I would welcome any views that you may have as to ways in which Automatic Train Control systems could best be optimised to create the greatest benefit for UK railways in both the short and long term?

Going back to basic principles, I still think that the single thing you could do that would be the greatest help would be to actually come up with some useful definitions of what Automatic Train Control is. It would have to be associated with ERTMS. The problem at the moment is that we have an ERTMS SRS which contains within it several things that hint at being train control, but really are only ATP and in cab signalling. I don't think that the specification has really begun to even think about train control. That is a long way off. I think it is important to start with a proper model of how the railway works. How is it that trains run safely and meet the operational need.
You need to start with timetabling. The timetable was the first defence against train classes, and is still a significant defence. A model then needs to be built up from that to show where the safety comes from and where the business efficiency comes from, and making sure that the right parts of the system deal with the right things. So, I think a train control system is a different thing from ATP.
You can have an ATP that only imposes certain constraints, whereas a train control system could be doing all sorts of things tied in with train scheduling and efficient train management. Deciding what speed the train should run to be the most fuel efficient. There are lots of things that you could learn from the aviation industry. I think at the moment that the railways are so far off that you can not just rush into these things. We talk about ERTMS as a trasin control system, but it isn't. It is an ATP and a cab signalling system unfortunately merged.
You need to think about what the boundary between a train control system and ATO. It needs to show what is ATP train protection functionality, what is train control (to what degree does this system support the efficient running of the train), ATO (a set of functionality which replaces the driver) or if you have a driver, what is his role? You need a model which defines all of these parts and makes it clear that they have different functions. At the moment the Rail industry does not have that view at all, and it is continually getting itself into scrapes by muddling the functions.

Qu 5. The terms 'Signalling' and 'train control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control system'?

I support the fact that they are seriously muddled.

Signalling is the set of functions that actually provide movement authorities to the train, whether to a driver or ATO. It maps on to the colour lights we have at the moment. They are the underlying sub-system that provides movement authorities. That would include speed boards and other fixed signals. The means of providing information to the entity responsible for driving the train.

Train control is the local control of a train, so it would include those aspects of advising the thing that drives the train about features that support efficient running. I would expect it to consider issues like timetable, fuel management, constraints on operation due to penalties for lateness. Things which in the past have been dealt with by the driver. It considers the train in isolation.

Railway control is up a level. How can we get the most traffic through the infrastructure.

Would you see Signalling a subset of train control, or as a separate thing?

I think it is best to handle them as separate. I would start with the concept of a driver (which may be an ATO), a thing which actually controls the locomotive power and brakes. At the moment we have drivers who fulfil the function of both train control and ATO. ATO executes the strategies of the train control function. You could argue that train control should be the higher function that sits above many other functions including signalling and ATO/driver, but then you have to come up with some view of how they fit together. You need to have something that has to meet all of the objectives of a train and decide strategy, it then issues instructions to the ATO or driver. In arriving at those strategies it will take in information from all sorts of different sources. With respect to safety, the biggest source is the signalling, but it may also take in train health, wind speed, the amount of fuel on-board, passengers on-board, all sorts of things. Train control then has the overall executive strategy of deciding how the train should run. ATO then has to implement that strategy. Signalling is a source of safety related input to that strategy

I think that there is a fundamental split between trains and track. Trains are discrete entities and the track is a distributed infrastructure. I think it is of enormous value to think of trains operating on track, not to think of them being joined. This view is in complete conflict with the rest of the rail industry's view. I think that the split between train and track is so important that it should be recognised. The better concept is to recognise that trains are objects that have a life of their own, maybe owned by different people with completely different management strategies. By that approach, you have to see train control as purely train-borne. I would not see it as having any part trackside. It is how the train decides its strategy for achieving all of its objectives.

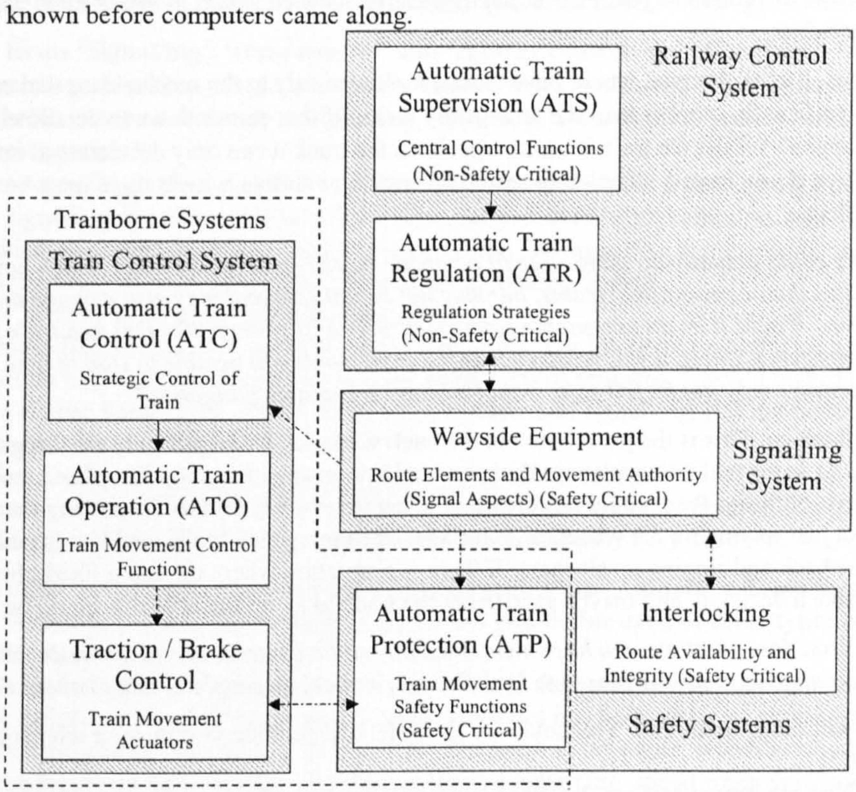
I would have ATP as a completely different thing because I have a mental model that I believe would work well, which is to split control and protection. The control system will be a complicated thing. It will have all of the business drivers. The ATP is much simpler. It has a very restricted view – is this train infringing any known safety rules. By apportioning responsibility between the two, you end up with one system checking the other – which is a strong defence. So, train control would be the strategic control of the train – on the train. It would provide advice to the driver as to how to achieve those strategies. If you want the driver to act as a human servo is that a good thing? Why is he there? At the moment a lot of train control functions are allocated to the driver. If there is no such thing as ATC, then all of the control functions and operation functions are allocated to the driver.

I see those as local strategic conditions with ATP in parallel with it, not a part of it. It is sitting beside it like a drivers mate. Train control definitely has ATO beneath it as that is how it implements its strategies. ATS functions are nothing to do with the train. They are inside the railway, to do with running multiple trains. Fundamentally the interaction between trains is the infrastructure. There is a different model that is now being used in air traffic control, which has each plane detecting the other planes around it and taking decisions on that basis. You could do something similar on railways, where each train is responsible for detecting all other trains around it, and ensuring that some separation is maintained. That would be a fairly radical step, but is by no means impossible. You could envisage a system where trains determine their position by GPS and transmit that in a broadcast fashion by GSM. All trains within say 10 km could receive that message and keep a model of where other trains are. Each train could then have responsibility for maintaining separation if they were coming up behind another train. That would be a dynamic model. You would have the problem of determining which track the train was on, but that is not impossible to imagine. At the moment it is probably a step too far, but it could be possible in the future.

At the moment the most important boundary is the train / track separation.

The interlocking should not be a part of signalling on the same basis as ATP should be separate from ATC. The railway has lost its way. Interlocking originally came along as a simple protection mechanism to prevent signalling errors. It was always meant as an add on to the signalling system, not a part of it. It monitored the signalling and actively prevented certain conditions that were unsafe. That fits my model of separating control and protection. If you kept the interlocking separated, off to one side but with the power of veto, the railway strategy could be determined by supervision and regulation instructing the signals directly. The interlocking would then monitor the instructions to see whether any of them go against any fundamental safety rules. It could again be a fairly simple thing that just checks the system and shuts it down if it goes wrong. The regulation type functions are then expected to operate safely and take on the first line safety role. Timetabling, local strategic route planning and the signalling as layers of that process, with the interlocking off to one side as a safety net.

At the moment railways seem to have missed lessons that other industries have learnt – and that they even seem to have known before computers came along.



Qu 6. The introduction of higher speed operation, higher capacities or changes to track layouts may all offer operational benefits, will influence the commercial viability of projects and will also affect safety. How do you think the trade off between safety and operational flexibility can best be managed?

In order to sort this sort of thing out, you need to decide your strategies early on. How is the railway going to work with your system.

For example, above 125 mph, how is the system supposed to work. What will the driving strategy be? How is the driver expected to work. Will it be head down driving or head up? You need to analyse whether you have something that is safe or not at that level before you get into designing the details. The ERTMS specification, for example, is not stable or proved, so you can not just say it will be safe because you are using it.

The big decisions on the trade off between safety and operational flexibility are taken earliest. They have to be supported by safety analysis before you kick of you design process. Otherwise you will find that you commit to options that will not be justifiable on safety grounds.

Qu 7. Do you believe that operation within relative braking distance is a feasible option for application to UK railways?

Generally trains can not stop instantaneously, but they can do. The big problem is that you are introducing a new risk. I think that it is certainly open to consideration and something that people should be thinking about. There is already a model for it in the way that the roads operate. In air traffic control there is no concept of braking, only avoidance. There is probably not much scope to do that on a railway. I think it would make a very interesting but difficult study to determine how you could convince people that it was worth doing.

I wrote a paper for the IRSE Younger Members and LUL Technicaql Society on this subject and concluded that one justification was to prove that your increased capacity would attract people away *from roads, and since rail is so much safer than road travel that would effectively reduce your transportation risk – If there was a political will to accept that kind of argument.*

I think that is right. You would probably not be able to justify it locally within the context of the railway, you would have to pull in wider benefits.

One idea within the context of the railway is that of crashworthiness. If you could prove that the train was crashworthy enough to take a collision at a certain speed without causing serious injuries to the passengers, then you could obtain capacity benefits by operation at a braking distance that would result in a collision at that speed if the first train did stop dead. That may justify increasing train crashworthiness to get better capacity benefits – which would increase safety for other types of possible accident.

Yes, I think that you would have to also look seriously at the mechanisms that exist that could stop trains significantly quicker than the emergency brake. If that comes down to derailment say, how often does that happen? Whilst the train is on its wheels on the track it can only decelerate at its maximum braking rate. Even if you have a completely seized up engine or sliding wheels there must be a limit to the braking effect.

We could actually do things like fit magnetic track brakes to the train, which have hugely better brake rates than conventional trains, but may not be fail safe and may also damage the track you are going over. Would it be an acceptable safety measure to fit systems like that to a train for use in a real emergency where it was realised that the train would not be able to stop in time. Considering this measure may not be fail safe, would it be an acceptable mitigation?

I think so. This is the parachute out the back scenario. There probably are things that you would not want to do in normal circumstances that you could do in an emergency. People on mountain railways use all sorts of things they would not dream of using unless they had to. Something that actually grips the rails, not just relying on the wheel rail interface. The emergency brake could be literally throwing an anchor out the back and ripping up sleepers. If there is a situation where there is a fixed obstruction, you have to make a decision, and maybe sacrificing the track to avoid the accident would be worth it.

As one last question, you have mentioned air traffic control several times, do you have any recommendations for places to look for information, or people to talk to about that?

I will think about that. You could try the Safety Regulation Group, their telephone number was 01293 56717.
There are some basic text books on air traffic control. There used to be a library at the civil aviation authority, SRG group at aviation house in Gatwick that was open to the public. It is reference only, but you may find that of use.
There is certainly a lot of work going on there that would be worth your thinking about. There is TCAS, a plane mounted collision avoidance system which could be a model. In areas of real air traffic control it is a supplementary system, but in other areas they do rely on it. There are also vast areas of Africa and South America where a frequency is allocated for 'self-help' air traffic control use. Pilots use that to call out their position and ask other pilots to report their positions. Then they all plan their routes co-operatively to avoid each other. It would be worth considering whether that would be a valuable approach for railways.

Sir David Davies CBE FREng FRS (President, Royal Academy of Engineering and Chairman of Railway Safety).

Questions Raised by D. Woodland on 07/06/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry
Following completion of a PhD at Birmingham University, joined the university staff. For part of that time also working part time as a Senior Principle Scientific Officer at the Royal Radar Establishment, Malvern.
In 1967 appointed Assistant Director of Research at British Railways, Derby with responsibilities for research in communications signalling and automation. This included the development of Automatic Train Protection and cab signalling systems.
In 1971 appointed Professor Electrical Engineering at University College London, also serving as Vice-provost for 2 years. Around this time also consulted briefly for British Rail, and in the late 70s joined the Research and Technology Committee of the British Rail Main Board, remaining on it until privatisation.

In 1988 became vice-chancellor of Loughborough University of Technology.
1993 to 1999 Chief Scientific Advisor to the Ministry of Defence.
1998 to 2000 Non-executive chairman of the Defence Evaluation and Research Agency.
Non-executive Director of Lattice Group Plc and ERA Technology Ltd, Chairman of the MOD Nuclear Research Advisory Council and a member of the Strategy Board of the BAE Systems Virtual University.
In 1995 produced a report on Automatic Train Protection for the railway, ot the request of the Deputy Prime Minister.
President of the Institution of Electrical Engineers 1994/5
President of The Royal Academy of Engineering from 1996-2001, Chairman of Railway Safety.
President of the International Council of Academies of Engineering and Technology Sciences 1997/8

Qu 2. The terms 'Signalling', 'train control' and 'railway control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system?

In the simplest sense 'Signalling' describes a method for preventing trains from running into each other. Having some signalling of course, it is actually used to perform some train control as well– controlling priorities at junctions for example. I would use the definition that signalling is a mechanistic process to prevent trains running into each other if they behave the signalling system.
'Train Control' is more to do with traffic management.
Train protection is like an extension to signalling in that sense. The interlocking performs a very important function to prevent the signalman making some sorts of mistakes. I would regard train protection as a similar thing for the driver. It is just an extension of signalling, it is not to do with train control unless you take it sufficiently far.
ERTMS levels 1 and 2 are just signalling systems as they stand, not control. Level 3 would be control and fluidisation is also control. The system on the Docklands light railway is also control. It would be worth your time speaking to them. There they are doing work which is clearly very different.

Qu.3 Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

I use the term supervision in the sense of speed supervision, in the same way as I use train protection. It is a more precise definition than train protection because it implies it is being done by supervising speed. There is an interesting issue, which I don't think that I discussed in my paper, which is whether you should ever give the driver an indication of recommended or maximum speed when you have an ATP? I suspect that ideally you shouldn't for the safest system. Obviously, once you give that information to the driver he is likely to drive within it deliberately. You are then depending on your system being 100% good. If you don't tell your driver anything, you have two independent systems collecting their data (one maybe looking at signals outside and the other collecting data by radio, or whatever means it is). For there to be a severe accident, both of those systems then have to fail simultaneously, which is exceeding likely except where there is a common failure (such as an interlocking failure).
'Regulation' is normally used in the area of deciding priorities of trains at junctions. It is a control function. It is about sorting out conflicts according to an algorithm or priorities. There was a lot of work done when I was at Derby on junction optimising algorithms. A lot of papers were published by Malcolm Savage. That work was originally done on simple combinations of junctions and traffic flows. Then it was applied to some real stations – Glasgow Central as an example. It was provided as advice for the regulators in the signal box who were handling disturbances. There were frequent disturbances there because the long distance night sleepers would arrive to terminate at the station and bottleneck over the bridge. They were followed by the commuter services. If there were any problems along the line, a couple of the sleepers might arrive an hour late. As they were long trains, they would take up 4 platform sections just at the time you wanted to run your morning rush hour. Someone then had to re-invent the morning schedule on line. This system could actually do that on line and present it as a recommendation. I don't know what value it had in capacity. It was used in other places too.
These sorts of things need to be done on a network wide basis, because if you decide to delay a train to avoid a later conflict, you may just be creating a new conflict elsewhere in the network. You therefore really need a network wide coverage of that sort of optimisation technique.

Qu 4. How would you define the capacity of a Railway system (and what would you consider to be appropriate units for your defined measure of capacity)?

I gave a lecture recently to the Railway Studies Association on train protection, and I was asked about the impact of train protection on capacity by an operator. Operators are pretty experienced at knowing that if they design a timetable, they can work out what the maximum capacity would be for a section of railway,

but they would not dare to actually run their trains to that capacity because a slight disturbance and the whole thing would lock up. So, the question is how near to theoretical capacities can you run. I believe that if we knew the answer to that, it would be a very useful measure of capacity. It could be measured as a % of theoretical capacity, and that would depend on the speed mix of trains (whether they all have the same acceleration or you are mixing freight and other trains). The really interesting thing would be how near to the theoretical capacity you could get, and what you would need to do to get nearer. We need to talk about network capacity, with the possible exception of fairly simple metro type lines where all trains have the same performance, stop at all stations and there are no crossing movements – then you can talk about plain line capacity. For most railways, we need to talk about network capacity, and the only thing that I can think of that would be useful to define is percentage of theoretical. First of all you have to have a demand. Then you have to try to meet that demand with stock that have specific speed constraints, and then you can do a pathing diagram to work out what the maximum capacity would be. Then you can see how near to that you can get.
The simple forms of train protection (as applies to the western region for example) where you have a point on the ground where you communicate the signal aspect, have been shown to reduce capacity. There is a very simple reason, the driver can see the signal maybe 400 yds away, but he does not get told about the signal aspect until he is nearer than that. So, he can not take note of the fact that although he was operating under the assumption it was a Y aspect, it has actually cleared to green. He has to follow the ATP indication, which means you get a reduction in capacity.
Equally, if you communicate the signal aspect to the train throughout the block section, way before you can see it visually, then you get an increase in capacity.
I have always found it disappointing that there has not been much published work on how you can improve railway stability. When I worked for BR research about 30 years ago we had done some research, some of which was published, on what we referred to as fluidisation. What we meant by that was, if you got beyond the problem of communicating the signal aspect over a complete line and then took a more strategic view, you could see conflicts approaching a junction ahead. You could then often better resolve the conflict by delaying a train to minimise it, rather than bringing a train to a complete halt. That saves energy and time. People did very simple simulations of a few trains approaching a junction to show that you could save time. No one to my knowledge has ever simulated a reasonably sized network. I think that it would be very interesting and useful if someone chose either a real network or a standardised network (to include the sort of things that occur) and then looked at different strategies to determine:

- what the theoretical figure would be for the network
- how near to that you can get with classical line side signals
- how much improvement you get from continuous cab signalling
- how much improvement you get from a bit of strategic fluidisation
- improvement you get from moving block

These are figures that I don't think we know. It is ridiculous that this work has been running now for 30 years, but no one has done these things that should be fairly easy to do.
There is one piece of work that I am aware of. At the time I was with BR, we had a linkage with Loughborough University. A few students did PhDs at Loughborough linked in with BR research. There was one called Eddy Gelbstein in 1973-5 ish. He was working in the railway technical centre and had a part time thesis.

Qu 5. From your perspective, what would you consider to be the main issues to think about when introducing an ATP system onto an operational railway?

If you are talking about adding it to existing stock, there are huge difficulties in fitting things to old stock. The public and the press have huge difficulties in understanding why it takes so long. There are certain types of train out there where they just can't find anywhere to fit an antenna.
Similarly, it is easy to say that you always put a speed trap just before a signal, but when you go to look at a signal you will find a crossover in just the wrong place, or the relays that you want to interface with will be somewhere else and you can't interface to them. It is those sorts of things that make it difficult to put things in quickly.

Qu 6. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

The most effective thing would be if you could demonstrate that you get some sort of increased capacity from it, either by something like continuous cab signalling or fluidisation control optimisation. Even with small amounts of capacity improvement, it will start paying for itself in a way that safety is difficult to justify. There are other benefits anyway. When I spoke to Sweden, they told me that when trains cleared a speed restriction, drivers almost invariably began to speed up as soon as the front of the train cleared it.

That meant that the back of the train went way above the speed limit. They very rarely had derailments, but they did get excessive wear of the rails. With ATP, that was prevented. They had no figures to suggest that these benefits of ATP were so important that they made a big difference to the cost of ATP, but the advantages and disadvantages certainly at least balanced.
Another thing to consider is whether there will be a benefit in terms of driver training. Will drivers need the same amount of route knowledge? That I think is a terribly difficult question. It goes back to what you tell the driver. If he uses his own route knowledge he is independent, if he is dependent on the route information from the ATP system he isn't. If there is then a mistake in the ATP system, such as a temporary speed restriction incorrectly fed in, then that is potentially quite a dangerous situation.

Qu 7. I would welcome any views that you may have as to ways in which Automatic Train Control systems could best be optimised to create the greatest benefit for UK railways in both the short and long term?

How do we get better capacity out of it. It is the same thing.
One thing I am often not sure about is whether we should actually communicate the information about track geometry ahead to the train as it goes along the track, or put all of that in a chip in the train, which contains the entire geometry of the entire network. Then you just need to communicate point setting information. Once you know the route setting ahead and the signal aspect, everything else is fixed information, except temporary speed restrictions. As you look at radio systems, which will be cheaper to install and maintain (although a lot of people are worried about radio coverage), I suspect that is the way to do it.

Studies both on the WCML and in Holland have shown that the communication system delays on ERTMS level 2 will mean that Level 2 offers less capacity than level 1. I believe it is largely due to call set up and processing delays.

I had heard something about that, but did not know why.

There have been suggestions as a result of the communications delays that it might be better to go for some kind of train based control, where you communicate as little as possible with the central control system, and each train decides for itself where it is going and how far / fast it can go.

I think that it true. I am a great believer in placing as little equipment at the side of the track as possible. You can't get at it there. It is not a nice environment, or one that you can maintain things in easily. The last thing that you want is possessions. You have to have them for track maintenance, but you don't need them for electronic maintenance. All of that stuff can be somewhere else miles away.
We are going intensively into train intensive rather than track intensive systems. You can maintain things on the train easily. The maintenance person can get on the train and check up on things as it is going along. Then at the next station he can pull a rack out and replace it if necessary.
I know there is a lot of debate over the effectiveness of radio coverage. Whether to have two separate networks and that sort of thing.

Qu 8. The introduction of higher speed operation, higher capacities or changes to track layouts may all offer operational benefits, will influence the commercial viability of projects and will also affect safety. How do you think the trade off between safety and operational flexibility can best be managed?

You talked about that In your paper quite well. The cost benefit analysis has been done for this. At what stage will the 'public' or the media accept some of these discussions? They keep saying 'you can't possibly…', but if you think back 15 or even 10 years, the to discuss the design of our national health service in a way that you accepted you sometimes got things wrong and people died, was just unacceptable. It has now become acceptable. It is not yet acceptable on the railways, but I think that it may become so. After Southall there was a huge outcry, Ladbroke, much bigger outcry, Hatfield less so. There were not many people killed there, and people started realising that there were other impacts from that sort of situation. After Selby there was no outcry at all, even though lots of people were killed. Maybe things are changing a bit. It will be interesting to see what happens when Cullen part 1 comes out, later this month I think. It is interesting that the public has become a bit more realistic and used to it in hospitals. They are much more outraged about people who died in hospital under perfectly good care who had some of their organs removed after they died than they are about people who nead not have died, but did die because of inefficient care. That is very peculiar. Our cancer survival rates are not good, they are near the worst in Europe, but people do not seem to be nearly as worried about that as they do about the removal of body parts.

Qu 9. What information do you think:

A. Is required by a driver in order to perform the tasks asked of him
B. Would be beneficial to a driver to assist him in the performance of his tasks

C. Would be detrimental to a drivers ability to perform the tasks required of him

We know what the driver needs to drive independently – the route knowledge information, signal aspect information, information about the route ahead and his own train.

Don't assume that although the ATP system is designed to eliminate the driver's human mistakes, that you remove all of that risk. Some of the risk is transferred to the person who writes the programme and puts the track data down, or the driver who gets into the locomotive and enters incorrect data on train lengths and braking performance. Mistakes there can still cause a problem. You may have dramatically constrained the risk, but you have not done away with it.

There was an example in Sweden some years ago of a miss-programmed balise.

One interesting thing there is that they don't have any ATP in low speed or station areas. Something like 90% of all train movements are covered by it, but only 60% of track miles. Station areas have a speed limit of something like 30mph, and that whole area does not have ATP. If you think about it, that would be the complicated area to apply it to. If you think about the complications of applying ATP to a complicated junction area, a train approaching the station needs to know all of the possible routes ahead. The computer needs to know which ones could be set, and its speed restriction may be a function of all of the routes that could be set. That concept of having to switch information to the train, and at what stage you tell it which route is set, is very complex. You don't want to tell it too soon, in case you change it.

Qu 10. How do you think that your answers to question 9 would change if the system included ATO?

I suspect that you would only allow manual driving under special precautionary rules into the next station to remove the train from service. So the driver would not need so much information.

Qu 11. I was fortunate enough to be able to attend your presentation 'How Safe Can We Make Our Railways' at The Royal Academy of Engineering last year, and I have subsequently read your report 'Automatic Train Protection For The Railway Network In Britain - A Study'. During your presentation you made a comment that is not reflected in either the written paper or the report, that you could see no reason why railways should not operate in the same way as roads - with less than braking distance separation between trains. Do you believe that operation within relative braking distance is a feasible option for application to UK railways?

I made a slightly off the cuff remark that maybe shouldn't be taken out of context about relative braking systems, but it is more or less what happened on the roads.

As an aside there have been discussions, papers and conferences about 'if only we could automate roads like we do railways, we could increase capacity'. You realise what a complete nonsense this is. The moment it was automated, the automation system, or the company who developed it, or the minister of transport would be at fault if anything went wrong with it. I think it will probably never happen, because you would get a reduction in capacity, not an increase, by the time you had reduced risks to a level that were acceptable.

I think that there is quite an interesting question that you can raise though, if you assume that it is very unlikely that you would get something less safe than a moving block system, like relative braking say, what are you worried about? You are worried about a vehicle hitting something else, and the vehicle behind then being unable to stop before it hits it. The interesting question to ask is how more unsafe is that than just having the original train twice as long. If it was twice as long, the whole train would be involved in the accident. It all gets braked to rest very, very quickly. Alternatively you may have only half of the train involved in the accident, maybe the front half, or it could be a side swipe, and then a second accident as the following train runs in to the back of it – but that would happen at a reduced speed. Is that likely to be more damaging than a train twice the length running into a wall, or less? I don't know the answer to that. So, if is an interesting point if people say it is unacceptable to have relative braking distances, to say well is it acceptable to have two carriages in a train, because the back carriage will be involved as well as the front carriage.

I have wondered whether if you could prove that a train was crashworthy enough to take a collision at a certain speed without causing serious injuries to the passengers, then you could obtain capacity benefits by operation at a braking distance that would result in a collision at that speed if the first train did stop dead. That may justify increasing train crashworthiness to get better capacity benefits – which would increase safety for other types of possible accident.

I that is an interesting question. I find it amazing that this discussion we are having must have been held hundreds of times over the last 20 or 30 years, but no one I am aware of has actually gone and tried to do the sums.

BR had a research department with a few hundred people for the past 30 years. I am told that an entire library of their reports is still held by AEA technology (rail) in Derby. I suggest that you try talking to them.
I might be able to find a few papers about work being done in Derby when I was there. I will try to send you copies.
Relative braking may be worth looking into, but before that we need to understand moving block. We don't know how good it is at the moment. People talk about moving block on a straight line railway. I don't think that we know how to run moving block on a real railway. I think there was a lot of debate about how to do it within Alstom for the original plans for WCML. When you have a junction, that is a fixed block, so suddenly you have to mix moving block and fixed block concepts when you come to junctions. I don't think that we really know how to do that.
Docklands light railway is moving block, and I believe that one of the German underground lines has been for some time. It would be worth your while finding out about them.

You mentioned comparisons between road and rail. Are there any other parallels from other industries that you think would be worth considering?

Well, air is the obvious example, and marine. Both are dominated by international rules, and your own national ones are subservient. They are not routed in the same way, although there still are routes. With trains approaching each other on the same route though, the probability of a collision is much higher than when two aircraft approach each other on the same route!

I have wondered whether there may be comparisons with computing and telecommunications from the capacity point of view.

A lot of academic work has been done on telecommunications and switching systems, routing traffic all over the world to cope with demand. A lot of queuing theory was developed for telephone switching. There may be some interest there.
I see that the EPSRC (Engineering critical science research council who fund research at universities) have put out a call for research in this area now.

Interview by e-mail with:

Markus Eberhardt (Project Engineer – Automatic Driverless Operation, Deutsche Bahn AG).

Questions Raised by D. Woodland on 17/04/2001. Responses given by M Eberhardt on 19/04/2001.

I recently attended the IRSE 'Future Trends in Signalling and Train Control' seminar in Birmingham, at which you presented a paper on 'Driverless Operation of Main Lines'. Having considered your presentation in more detail, I have identified a number of areas for which I would be interested in more information:

Qu 1. Your presentation has been published in slide form only. Do you have a written paper on this or any related topics? (The short statement nature of slides makes their content often open to interpretation and can make understanding the true meaning difficult).
I'm sorry, but there is no written (long version) of the presentation and our own material is written in German only. But I think, I can help you in some points.

Qu 2. You referred to an ATC system (AFB) during the presentation. The term ATC is usually applied to any combination of ATO (train acceleration, braking, station stops, etc.), ATR (regulation) and ATP (the safety critical protection systems), and under that definition is not an entity in itself. However, it is clear within your presentation that the ATC you refer to is a separate sub-system. I would therefore be very interested to know how you define ATC?
You are right, there is some confusion about the definition of these terms, and therefore I used a slide to give the definitions use in the presentation: The definition you are looking for is ATC = ATO + ATP (+ATS), that I think is used in English literature. While preparing the presentation I found different definitions in several papers, so I took finally the one you found in the slides: ATO = ATC + ATP. But I think, today I would prefer the "English" definition to reduce confusion.

Qu 3. I am also not familiar with the acronym AFB. I would therefore be very interested to know about this system, what it does and how it does it.
The AFB (= Automatische Fahr- und Bremssteuerungsystem) is in some ways comparable to the autopilot- system in aircrafts. In Germany, the AFB is used with the LZB (=Linienförmige Zugbeeinflussung) that is a train surveillance and train guidance system used for lines that are operated

with more than 160 km/h. The train guidance information 'allowed maximum speed at this track section' of the LZB system is used as input for the AFB system, that controls acceleration and braking of the train to approximate the allowed maximum speed. In my presentation I refered to AFB as a technical solution of the function ATC.

Qu 4. In Slide 11, you referred to 'safety aspects of the automation of the functions of train control', dividing sub systems into safety-critical and non-safety-critical. I would be very interested to know on what basis this apportionment has been made.

At the actual stage of the project, this apportionment was done by the project manager. Finally, there will be a system hazard identification based on CENELEC 50126 to confirm or update this apportionment. So this is an iterative process that includes a lot of discussion with the EBA.

Qu 5. From slide 12, I would be very interested in details of the risk analysis that has been carried out to compare the train driver function with automated system functions.

There will be a risk analysis for the complete system of automatic driverless operation. We now started with the risk analysis of the obstacle detection. The basic principle is the comparison of the train driver with the technical system. The analysis is based on the German accident statistic. The number of accidents per year is divided with the criticality of the hazard to get number of hazards per year. (The criticality is the result of the modelled behaviour of the train driver.) The number of hazards per year will be multiplied with the criticality of the technical system to get the number of expected accidents, if the technical system is used. (The criticality of the technical system is modelled based on the specification of this system). Now the number of expected accidents for the technical system shall be smaller than the number of accidents in the statistic for the drain driver. If this can not be archived, the effects of other subsystems will be taken into account.

Qu 6. I would be interested in details of the obstacle detection and collision detection systems referred to within the presentation.

The obstacle detection is developed by several companies in competition. There are two main principles. The first kind of systems is based on video picture processing, the second kind of systems is based on radar picture processing. But there is no system ready to use in railways application. The systems used in car traffic are not suitable because of the requirement for a longer range of the system because of the longer braking distances.

The collision detection is not specified yet.

Qu 7. In addition to this, I would welcome any suggestions that you may have for other published material on these subjects that it would be useful for me to read, or for other experts whom you would recommend that I contact about these subjects.

Most publications about driverless operation are on subway systems. But the people that collected publications are in holiday.

Interview by e-mail with:

David Fenner (Engineering Manager – Train Protection, Railtrack Assurance and Safety).

Questions Raised by D. Woodland on 24/01/2001. Responses given by D Fenner on 25/01/2001.

Qu 1. I was wondering what figures for cost and effectiveness were used by Railtrack when they assessed the TPWS system for use?

Firstly you will find a lot of our data and Dr. Andrew Evans data similar since he has based his statistical analysis on work originally done when TPWS was first developed.

As part of the TPWS study AEA surveyed every fatal accident since some date around 1970. Fatal accidents were chosen because evidence and data tends to be more complete and robust than "serious incidents". Thus the data set is reasonably complete (but not absolutely complete).

The accidents were then categorised as ATP preventable or not. i.e. SPAD and overspeed events were considered to be ATP preventable adhesion related events or plain derailments were not. The ATP preventable accidents were then investigated and as far as possible the approach speed determined (from investigation of the accident report) as was the overrun distance before the accident and other relevant parameters.

At the time of the report it was expected that TPWS overspeed sensor would be placed at 200metres before the signal and the train stop at the signal. Then based on an emergency braking curve (which may have assumed 12%g) it was calculated whether the train would have stopped before the collision point or not due to TPWS intervention. If the train would have stopped the accident was counted as TPWS preventable. If the train would have slowed more than the driver slowed the train then the consequence in

terms of "equivalent fatalities" was reduced pro-rata to the reduction in speed. Separate evidence suggests that fatalities and injuries over a reasonable sample are most closely related to speed (rather than speed squared).
Overall the report concluded that TPWS would reduce "equivalent fatalities" by 68% but emphasised the method and volume of data made this approximate.
This work was then examined by Sir David Davies and using the same philosophy reworked and confirmed similar results. Atkins also redid the work for (I think) the HSE. This was available to the Cullen inquiry and may be available in witness statements. Our work is buried in archive and I cannot quickly put my hands on it but you've seen the essence in Dr Evans reports/work. There was some dispute at the Cullen inquiry where other people presenting similar data from a different statistical start point arrived at figures from 35% effectiveness to 75%. The problem is the sample is thankfully small and thus you can quite easily manipulate the statistics. To get the answer you want.
As you may now be aware TPWS overspeed loops are now being placed 350 metres before the signal. This will enhance the capability but not by a great deal since braking distances expand quickly above 70 mph.

Qu 2. I would like to know your personal views as to the expected effectiveness of TPWS and whether you think that it can be seen as achieving the optimal ATP solution in terms of cost / benefit, along with an explanation of why you hold the views that you do?

The equation for optimisation is complex. Originally Railtrack intended to fit about 15% - 20% of signals, with all trains (except perhaps shunters and the like confined to short distance movements) fitted. Given that the historic equivalent fatality rate is about 4 ATP preventable equivalent fatalities per year and even given the one in 100 year catastrophe does not exceed 6 then the cost was around ?4M per life saved.
In practice the Regulations require us to install at over 45% of signals at an industry cost in excess of £400M. Given a 20 year life and say a saving of 3 lives per annum then the ?400M is saving 60 lives at a cost of just under £7M per life saved. Obviously if you use 2 lives per annum the figure rises to £10M. However if traffic does increase as predicted then the accident risk and consequence will presumably also go up reducing this figure. This is all assuming discounting due to time value of money.
Here is of course the need to consider the press and public response to these accidents which are now considered to be preventable and it may well be that even more is necessary.
On a more practical note the vast majority of SPADs occur at far less than line speed either because the train cannot achieve line speed (stopping trains) or because the SPAD is a start away. Thus my essential belief is that many accidents as a result of SPADs will be avoided. And many of these will be potentially serious ones. Both Cowden and Ladbroke Grove would have been prevented by TPWS. However the isolated high speed SPAD such as Southall and Watford (marginal since the precise point the driver started braking is unknown) would not be avoided by TPWS alone. Southall would have been avoided by AWS since the driver would have be alerted to the double yellow. In my experience these high speed accidents occur about every 10 years but it may be getting worse. It is these very rare high speed accidents that will cause the deaths even with TPWS.
Railtrack, and I believe the industry, now accept that we must move to a full ATP system and that increasing capacity will make this more essential. The issue is to get benefit from the application. TPWS offers some potential benefits in terms of less disruption due to SPAD incidents but it will not get rid of all of them.

Qu 3. As an additional point to this, I would welcome any views that you may have as to ways in which ATP could best be optimised, what forms of optimisation would create the greatest benefit to UK railways and what ATP developments would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

ATP should get rid of all except adhesion related events. These impact on the optimisation of ATP but at present we do not have measurable data since it has not been historically collected. ATP will also allow us to simplify signalling (e.g. remove approach control) and ultimately at level 2 or 3 reduce the volume of trackside signalling infrastructure. Again these affect the optimisation equation.
Optimising ATP would ideally require us to install ATP as Railtrack re-signals and for the TOCs to fit as they replace trains. However this runs into human factors issues of partial equipment for a long period of time and requires a modified approach so that this can be managed. This is work Railtrack and the TOCs are just starting and will probably not be available for your work. However updates may be available over coming months.

Interview by e-mail with:

Mark Glover BSc ,BEng, Eur Ing, CEng, MIEE, MIRSE (Sales & Marketing Manager Westinghouse Rail Systems Ltd).

Questions Raised by D. Woodland on 05/03/2003. Responses given by M Glover on 05/03/2003.

Qu 1. I was wondering if you knew the frequency with which it was proposed that trains would transmit their location signal in the Westinghouse moving block system proposed for the Jubilee Line Extension?

Basically what we found in the past was that the update time for location is in fact the reason that Moving Block doesn't necessarily offer the advantages over Distance to Go ATP systems that you might expect!!
The JLE system would have had up to 12 trains communicating with one Moving Block Processor. By the time that you've set up the communication, passed the message and received the handshake, you're not where you were. Assuming that the worst case is probably as much as 1 - 2 seconds, at 80km/h you could well be nearly 45m out - move onto mainlines and 200km/h and
you're over 100m out!
Some of the systems that are out there – particularly in the states - use Spread Spectrum systems – again the latency can be an issue by the time you've mixed and merged the information from different radio channels.
Basically it's all down to getting the correct levels of security over a link - get this transmission time down, and you'll improve headway. Without manageable radio delays you are much better off having a simpler, Distance to Go, system.
Of course this is all without even thinking about other things like whether the positional information reported is accurate - allowing for slip / slide, distance since last positional reference and so on!!

Eddy Goddard CEng, FIRSE (Train Systems Engineer, London Underground Ltd).

Questions Raised by D. Woodland on02/07/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I began working on the introduction of Victoria Line Signalling in 1963. I have worked in all areas of signalling – maintenance, installation, design and development. I am currently responsible for rolling stock and signalling, with experience of moving block systems in Canada (a successful introduction) and the Jubilee Line (unsuccessful introduction). Also ATO systems world wide, acting as a consultant in systems – particularly in the USA and Canada. I am also an ex-president of the IRSE

Qu 2. The terms 'Signalling', 'train control' and 'railway control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system

Signalling is the safe enforcement of train movement and the direction of the driver. The signalling system should itself be divided into vital and non-vital. The vital part should be kept as simple as possible and capable of meeting a SIL4 degree of safety. In addition to that, there is a need for the movement authorities to be given to trains in a non-vital sense. I would see that as at highest SIL2. In both of the parts, very structured production is implied.
Train control is the efficient movement of trains to meet the service requirements. The overall function of optimisation of the train service. It has a lower degree of integrity. In modern systems, train control also goes on-board the train. There is then a similar split between vital and non-vital onboard systems. The vital function is a combination of signalling and rolling stock. The less vital function is the actual ATO type operation of the train.
Railway control is more about management control systems and general control systems. It is the overall operation of the railway to ensure safety and economic movement. Railway Control Systems also include trainborne functions, such as the passenger information displays. They can be of standard IT integrity, and rapidly changed.
Signalling is a part of train control, and the interlocking is a part of the signalling.

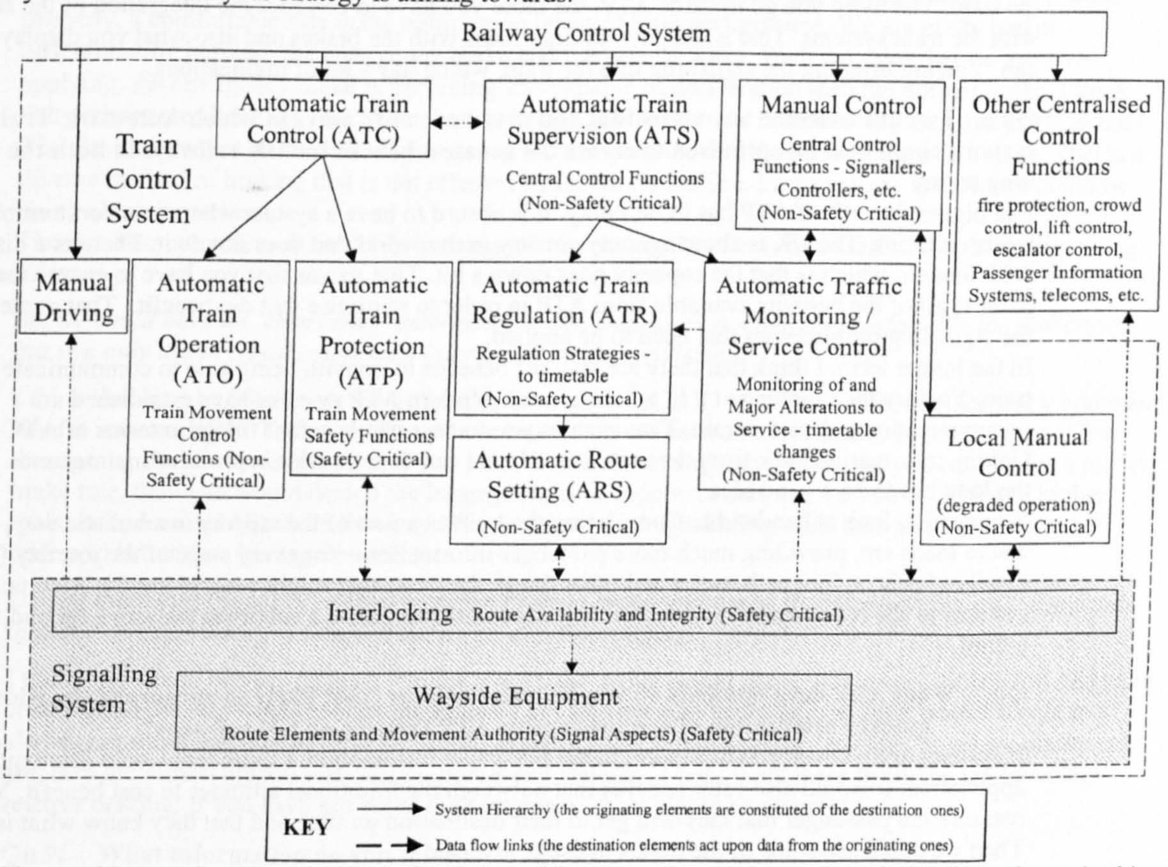
Qu.3 Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

Automatic Train Regulation I see as the giving of authority for a train to proceed based upon the timetable and location of other trains, without the need for manual intervention. That can vary from the crude approach of first come first served through a junction, through to sophisticated regulation algorithms to optimise the service.

Automatic train supervision I see as being inextricably linked with control rooms. It is about knowing where all of the trains are without having to do any manual operations to find out. So, track circuit information, point information and identification of trains all come back. Logging of the operation. Possibly also simulators to help replay incidents and for training. It can also include commercial aspects like the total delays to trains and the causes of those delays.

So, what is the relationship between train supervision, train regulation and train control?

Train control sits at the top, saying what it is that we are trying to achieve. Train supervision then sits below that, making sure that you know what is going on. Train regulation then sits below that, as the way you achieve what your strategy is aiming towards.



You could expand on the other centralised control functions. In addition to the relationships shown in this diagram, you also need to consider how the ATP onboard relates to the braking system, etc.

When you get down to functionality, people don't really differ in their views, but they do in the terminology that they use. We are currently working on a glossary in the IRSE Technical Committee to help with this problem. The Germans have very precise definitions of each thing, but translating then to English is not straightforward.

Qu 4. How would you define the capacity of a Railway system (and what would you consider to be appropriate units for your defined measure of capacity)?

If you consider pure metro, it is very simple – the journey time capability. We now have a specific measure that says how many people you can move from A to B, and as a measure of its quality how long it will take to get there, weighted according to stages in the journey. Waiting time counts higher, and variability counts even higher. We are then looking for the best performance you can get, the actual performance you can get and the difference between the two. This approach has been published in the PPP documentation. I also believe that Phil McKenna described it in his lecture to the IEE summer school.

To answer in a signalling sense. It isn't the bare minimum headway. You have to allow for a recovery time, to ensure that whatever happens the service will normally be stable – and recover over time.

If I were not from a metro environment, I would be talking in terms of paths. If you have mixed stock, you have to allow for the fact that a goods train will precede slower, brake slower and need more space than a high speed train.

Qu 5. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

The main issue if you have never had ATP, is to look again at your philosophy of signalling and operation. It does completely transform that philosophy and has profound implications as far as driver training, rule books and the layout of your signalling are concerned. Having ATP limits, but does not eliminate the danger associated with a SPAD. It also creates a whole new set of risks. As far as control rooms are concerned, it also changes things. It is a very profound change.

If you already have any form of ATP and move to another form, that again has implications. For instance, with introducing ATO and coded tracks instead of trainstops, the biggest implication is the process of converting the railway. As far as ERTMS or LZB or any other standard system is concerned, it is less of a change to make. You do still have philosophical differences though. When you think about cross acceptance of equipment, it is not so much the physical equipment that you have to think about, but the environment that it is used in. The social, human environment. It is really quite profound.

The other thing that you have to look at is the trains. In old fashioned non-ATP, signalling really ignores the train. The more you go towards ATP, the more you have to consider the integration of the signalling with the train systems. That is the physical interfaces with the brakes and also what you display in the cab. You are bringing the signalling into the train, which has a lot of implications.

Qu 6. I would welcome any views that you may have as to ways in which Automatic Train Control systems could best be optimised to create the greatest benefit for UK railways in both the short and long term?

The biggest benefit of ATP has to be safety. It is absurd to have a system where you don't enforce red lights on trains. The UK is about the only country in the world that does not do it. There is a dis-benefit of ATP though, which is that the capacity goes down a bit. That means that you have to ensure that you harness all of the benefits available from ATP in order to minimise that dis-benefit. That comes back to the signalling philosophies that need to be applied.

In the longer term, I think that there are spin off benefits to do with being able to communicate with the train. You are then looking at it in a holistic way. Modern ATP systems have established a communication link to the train. That enables you to do a whole raft of things in terms of ATC and ATS. Getting information back from the train as well, you can then do a lot in terms of maintenance. Closing the loop has to be a good idea.

You need to look at bandwidth. Considering the train as a part of the railway in a holistic way, knowing where loads are, providing much more passenger information – for every stage of the journey (with details of delays, foot path routes and other things the passenger might want to know). Also tailoring systems to the requirement. A rural railway has one requirement, a suburban railway a second, and metros a third.

Qu 7. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK railways?

If systems were developed to support better passenger information, I think that would drive quicker application. It would also offer benefits that outweigh the traditional attitudes to cost benefit. You need to reassure the passenger that they will get to their destination on time and that they know what is going on. Then we could see some of the trends in usage reversed.

Qu 8. The introduction of higher speed operation, higher capacities or changes to track layouts may all offer operational benefits, will influence the commercial viability of projects and will also affect safety. How do you think the trade off between safety and operational flexibility can best be managed?

I think that there is not really a trade off. The operational flexibility should be looked at as integral to the system. You need to define your system to make sure that you meet the requirement. The way that the safety system enforces its safety – brake curve monitoring, any of the distance to go calculations – you need to ensure that you have accounted for the highest speed you are going to get, but also that your train design reflects what is required. You need to take the vehicle manufacturer, signalling and passenger requirements into account simultaneously and design around that. The biggest problem I see at the moment is that the braking system and the motor system are designed in one corner, and the signalling system in another. You then get on one ATO train after another where you can feel that we are not designing as an integrated system. If you optimise that interface, then you are getting the best that you can out of the train. Then I think that the operational flexibility would follow.

Our statistics show that an ATO train is a much more efficient train than a manual one. On the Central Line now, we are getting minutes more recovery time than we were getting with manual operation. You get much more consistency with ATO. With ATP you are at a half way house. It all depends entirely on you're your human machine interface, and how well you are enabling the driver to perform in an optimal way. If you go to somewhere like Hammersley Mines, where they have 2 mile long trains, they have an ATO that is very different to anything we would have. You need to keep the forces on the vehicles constant. If you are going over a hill, you therefore have to control both the front and back train to keep the tensions right. That would be very complicated to do manually.

I believe that if you work out what you are trying to achieve and then make the whole system reflect it, you will get the operational flexibility because you designed it in at the beginning. It is a matter of looking at one control system not five!

It occurs to me that there are things that you could do to improve the systems performance. For instance, if you wanted to run trains faster but are worried about sufficient braking distance, why not improve the brakes? There are certainly things that you could do to improve the braking performance, but you might need to introduce things like jerk limiting, which would tend not to be fail safe. What are your views on that approach?

I think that you need to differentiate between stopping a crash and having a comfortable ride. If done properly, a comfortable ride is the optimum in terms of train performance. We are pretty bad at differentiating between the emergency brake and the normal brake. With the rules that we are now applying, the emergency brake is becoming less capable of deceleration than the normal brake. I think that perhaps we should be looking at diverse braking. What can we do about a system that starts to fail? The classic is leaves on the line. If you get on a tram, you find that they have magnetic brakes. That is a diverse method of braking that is not effected by leaves on the line. Perhaps there is more we could be doing along those lines.
There is also more that we could do in terms of minimising the effects of a crash through our signalling rules.

So, we could have an 'emergency' emergency brake, that might not be very comfortable for passengers, but you only use in a real emergency in order to avoid a collision.

I think that is right. You know the difference between what LUL considers a normal emergency brake and what it is on an HST? The only difference is that we do not have a hot tea trolley hurtling around! But then, going back to system design, if you knew that under emergency conditions you could have a higher brake rate, then you would design the luggage racks to cope with that, etc. So you could make higher brake rates reasonably safe in that way. The fact that you might end up with flats and a few burnt out brakes id better than having an accident.

Do you think that the use of operation within relative braking distance could be justified on the grounds that any collision would not exceed the trains crashworthiness?

I think that would be very difficult. On the whole, the driver would perish – it is very difficult to design in crashworthiness that would protect the driver. I am also not sure that public opinion would accept it. I also do not think that relative braking would actually achieve much. When you look at where bottlenecks are, they are almost always stationary – points and stations. Then I think there would be no advantage to relative braking. If you have not cured the bottleneck, there would be no point.

Qu 9. What information do you think:
- **A. Is required by a driver in order to perform the tasks asked of him**
- **B. Would be beneficial to a driver to assist him in the performance of his tasks**
- **C. Would be detrimental to a drivers ability to perform the tasks required of him**

It is very difficult to know this. At one extreme, you can expect the driver to have detailed route knowledge. I think that is no longer sensible. More and more, we can now provide detailed information to the driver to enable him to at least have reminders of the situation. You do need to let the driver know the speed that he should be going at and the distance to the next speed change. You can even tell him about the speed after that.
I am an advocate of ATO, because once you have the information available you may as well automate the driving and change the drivers role.
It is important that the driver has no surprises. You have to way up the fact that you are expecting the driver to drive the train to signals with the fact that he also has to avoid hitting anything in front. That is, trackside safety and obstructions. So, you need to look carefully at how much time you expect the driver to spend staring out of the front.
If you consider a classic metro with platform train interface problems and in-cab CCTV, you need to consider what environment you are aiming for and how you are going to enable the driver to prioritise his tasks. There is a danger of being over clever. You don't need to start switching displays automatically. I have seen an awful display that changed like that. The two screen s looked the same, but in one case it was the actual speed and in the other a target speed. There was a good engineering argument for doing that (to do with modes of operation), but there is a big potential for confusion. I think that you need to consider human factors and consult the experts who really understand the issues. It is not just about displays, but the whole environment. Both dynamic and static information, training and notices, clues on the wayside to tell them where they are. Adding all of that together, a proper human factors study with professionals is needed.

So, the sort of things that are detrimental are confusion, overload, unclear instructions as to their true purpose, bas training, bad recruitment, bad supervision.

I have often wondered weather railways have ever considered using head up displays?

Yes. There was a trial somewhere in Europe, but I can't remember where. I think that there is a lot to be said for it, although there are problems with head up displays. They tend to fix the position of the driver, where drivers (particularly on main line) like to be able to stand up or sit down, or move around a bit. On metros we have the same sort of thing where they have to operate doors and so on. There is a reluctance by drivers to accept that sort of thing. There are also problems with reliability – but if you can overcome that in a fighter plane, surely you can on a train. I think that they will eventually come along, because they are a good solution where you expect the driver to look ahead but also look at a display.

You said that it was important to give the driver a different role with ATP. What do you think that role should be ?

I think the train captain idea is a very good one. So that the driver is there principally to rescue the situation when something has really gone wrong. You could say that you don't need a driver at all, but that would be a bit fraught when you considered all of the potential hazards. On the other hand, expecting someone to do an entirely repetitive task all of the time would be rather boring. The kind of person who can do that tends to not be the best. With ATO, you have de-skilled the job a lot.
Especially looking at high speed railways, what could a driver in the front of a train actually do? At 200kph he is not going to be able to stop the train if he sees an obstruction. The warning time he can give is minimal. So, why is he in the front? I think that the customer facing role is the right answer for what the driver can do.
In an enclosed environment, driver-less operation is perfectly feasible and you do not see a bad reaction from passengers.

Do you ever have problems on the Central Line from mixing ATO and manually driven trains?

Only that the coded manual train is slower. The big dilemma is always do you fit all of the trains first and then do the signalling, or do the signalling first and then fit the trains. I think actually you can do both, but you must design the system around the situation you chose, not design it in afterwards. You need to look at that from a capacity point of view. It is not just what would be the effect on the signalling, but what would be the effect on the timetable, what are you going to tell the regulators and so on. You get things like should you fleet the ATO or ATP trains and then allow a couple of unfitted trains to follow them. If you can fleet the trains you benefit more from the fitment. What you can do , of course, is install extra signals for the unfitted train that you then remove once everything is fitted. That may be more expensive up front, but is worth considering. Also, would you ever end up with all trains fitted? If you think about West Coast for example, where you have minor services running across the main line, the cost of fitting all of the trains would be very high.

If we did have ATO with a train captain, he will not have any route knowledge when he does have to drive.

I think that you would need to make sure that every driver did drive every route on a regular basis, so that t hey are prepared for the emergency situation. It wouldn't need to be every day, but regularly.
If you assume that drivers have route knowledge, that creates a whole solution set. If you assume they have not got route knowledge, that creates a different solution set. Then you need to include lineside and on-board information, possibly speed governing (limiting the speed to line of sight driving). You would then need to provide bolt holes to remove a train from service, but what you shouldn't do is allow an ATO failure to stop your whole service. So, multiple scenarios, look at lots of potential solutions and select the best. Don't just assume there is only one solution.

Eddie Gosling Fellow PWI (Retired).

Questions Raised by D. Woodland on 08/04/2002.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?
I worked in p-way all of my career, starting as a lad in the drawing office and working my way through area civil engineer and p-way production engineer. Most of my time was spent on the South Western zone, with 750V third rail power supplies – so I am not very familiar with overhead systems. I am now retired.

Qu 2. How are line speeds determined?

Line speed limits are largely comfort limits rather than physical limits to safety. I am not sure anyone knows exactly how they have all been set. A lot of them go back into the annals of time. The route speeds have evolved, based initially on the geography of the route. The line speed has been slowly increased over time. On straight track you can do what you like. The main limiting factor is then the capability of the traction.

Qu 3. How are permanent Speed Restrictions determined?

You can also get specific speed restrictions to particular types of rolling stock due to their characteristics. One aspect affecting that would be the structure guage. There is also structural strength to consider - the faster trains go, the more vibration they cause. That can cause fatigue to worsen, so speed restrictions are imposed over weak structures.

On curved sections, through stations or at certain other features the line speed will need to be restricted by a permanent speed restriction.

On curves, the main factor to consider is the rate of change of Cant. If you have mixed traffic running on the line you can not provide the ideal cant for all trains. You can not provide the ideal cant for the maximum speed trains because the slower ones would fall off the side half way around the curve! You have to go for a medium value. There are a whole series of rules used to determine what cant is needed to go around curves of a particular radius of curvature at different speeds. The main factor considered is the comfort of passengers.

I have never seen anything written down that explains how the particular speed values used are calculated.

There are also restrictions due to the signal spacing and sometimes the power supply. That can even cause temporary restrictions from time to time if a sub-station is lost. The power available is then insufficient to allow full train speeds and a restriction will be imposed.

Then, of course, there are junction speeds – depending on the design of the S&C.

There can also be vertical curve considerations, but it is fairly rare for them to be significant.

Qu 4. What about temporary and emergency restrictions?

There are a million and one things that can happen to cause a speed restriction to be imposed. If there is a crack in the rail, a broken rail or a bit missing out of a rail, things like that, there are a whole set of speed limits defined in Railtrack Group standards. The man on the ground would determine the speed restriction to impose based on those standards.

There are various other reasons for imposing restrictions. Weather conditions, such as the likelihood of track buckles in hot weather, can lead to blanket restrictions. Again the speed values to use are pretty well defined in standards, although there may be some discussion in the area hierarchy to decide on the exact number if it is a hot weather restriction.

Maintenance work can lead to speed restrictions. For example, if you have just tamped the track you may need to apply local restrictions in hot weather until the track has settled back down. Also overruns of planned work can lead to a restriction being maintained for longer than it was supposed to. Then you just maintain an existing temporary speed restriction but it has to be treated as an emergency restriction because it is not as notified.

Qu 5. Can you suggest any contacts for further information about these subjects?

You could try:

Alan Green, the Chairman of my PWI section;

Bob Hazel, the p-way engineer at Waterloo;

Quentin Philips, Principle Track Standards Engineer with Railtrack.

Neil Harwood (Train System Engineer, Alstom Transport).

Questions Raised by D. Woodland on 16/10/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I started with BR in 1960 as an apprentice in the BR wagon works at Shildon. In 1969 I moved to the Railway Technical Centre at Derby to work on the testing and performance section there. I then moved from testing to design (principally bogie suspension design). I stayed there until 1989, when I moved to GEC at Manchester as their Chief Mechanical Engineer. Then in 1993 I moved to Birmingham as the GEC-Alstom Chief Mechanical Engineer.

For the last 3 years I have worked predominantly on the West Coast project as Train System Engineer.

Qu 2. What braking rates are achievable by use of different types of train braking system currently in use, either independently or in combination?

There are two questions there. Brake rates are all about economy – how sophisticated you make the braking system. Higher brake rates could be achieved with brake rates by putting more energy into disk brakes, but then you would come across the serious limitation of adhesion. There is also the issue of comfort – what retardation rates would you want to subject passengers to. It works out that the limits of adhesion and comfort work out with the economy to give a maximum of about 9%G for a passenger train. On older coaching stock, it was 7% based purely on block brakes. That accounted for some brakes being failed on the train and losses due to mechanical linkages. Accounting for all of that, you could practically achieve 7% day to day (assuming good adhesion levels).
In the days of BR, discussions would go on between the rolling stock engineers and signalling engineers to develop what was known as a 'W' curve. It was based around the way a typical train would brake based upon a minimum amount of maintenance. The signalling system would then be laid out in accordance with the 'W' curve.
These days, similar discussion goes on in most administrations between the signalling (or control) engineers and rolling stock engineers. Metro's would have different values than main lines, but they are all about 10%G for emergency and 9%G for service braking.
On EMU's using the West Coast today, the service brake rate is specified at 9% for speeds up to 200kph. At higher speeds, the brake is specified as 6% (until the speed comes below 200kph). That is largely constrained by adhesion considerations and to some extent by energy inputs, but mainly by adhesion. The emergency rate for a DMU or EMU up to 5 cars long has to achieve 12%G according to the Group Standards. Longer formation trains have to achieve 9% up to 200kph, 6% above that on the emergency brake as well.

Do you know which group standard that is?

I could find out, but I do not remember.

Why is there a higher specified brake rate for shorter trains?

Shorter trains are more likely to be involved in SPADs than longer trains. In poor adhesion conditions, if the first few axles of the train begin to slide and you engage WSP, you have lost a significant proportion of the trains brakes. If you have a 9 car train and the first few axles of the train begin to slide, the loss of braking effort due to WSP has proportionately less effect.

How certain are we that the braking rates we currently assume are actually achievable (and under what environmental conditions)?

A great deal of testing goes on. Stopping tests under dry rail conditions. There is no performance specified for low adhesion conditions. If the adhesion level became 0%, you would not be able to stop. You need sufficient adhesion to support the braking rate. There is usually a 25% contingency between the signal spacing and trains theoretical stopping performance. That is there to allow for a small amount of over speeding by drivers and for poor adhesion performance.
In conventional signalling and TCS type ATP systems, you rely on that extra contingency for the case of poor adhesion. With moving block you know where the other trains are, so you then need to decide what contingencies you are going to put in to the braking separation to allow for poor adhesion. The contingency allowed for today in the signalling system is there for good reasons. You will still need some level of contingency with new systems.

I don't think that issue is very well understood at the moment.

It never is. In my experience, most people think that adhesion is always there. There is only one person who can guarantee that it will be, and he is not terrestrial!

How much would it cost to bring all trains on the network up to the highest level of braking performance?

It would not be feasible to do it purely on the basis of cost. It would not be possible to justify it.
You would probably benefit from having a discussion with a brakes supplier.

Who provides the brakes for the new West Coast trains?

Knorr Bremse. I suggest that you speak to Jonathan Padison who works for them in Munich. He did a PhD at Loughborough University with Roger Goodall. I am sure that Roger would be able to put you in touch.

Qu 3. Are there systems that could achieve higher brake rates that are currently used elsewhere in the world or on other transport modes?

Whether you looked at Japan or France, it would be unlikely that you would find much difference. People try to maximise their electric brakes, either by regenerating the energy or using braking resistors. They try to use to friction brake as a last resort, because friction brake maintenance is extremely high – a major part of the maintenance costs of a train.

What about track brakes or eddy current brakes?

The Germans use magnetic track brakes. It is mandatory on most of their rolling stock. They have very short signalling distances, so they don't want to be totally dependent on an adhesion based railway. They use the track brakes to get more friction. They also get higher braking rates by doing that.
If people were laying out a railway, they would not want to use track brakes though.
Generally, no one gets more than about 12%G from conventional braking systems. Even then, that can only really be used for emergency braking. You put a lot of energy into the disk brakes which makes them very hot. You then have to stop whilst they cool down. You couldn't move off and do a repeated stop.
I am not familiar with eddy current brakes.

What would higher brake rates to the trains and track in terms of wear?

If you went to excessively high brake rates, you would come into a discussion with the safety authorities about what you are doing to the passengers – would luggage start to move around inside the train, would catering staff be safe cooking hot food, would people trying to move about the train in corridors be made unstable, would catering trolleys fly around and hit people.
There is a fine balance between comfort and safety. You need to consider these operating issues. They would be of more concern than wear. You need to consider comfort, safety, adhesion and the logistics / costs of extra equipment.

Qu 4. What application delays are applicable to different types of train braking system?
Modern EP assisted brakes (Electro-Pneumatic – using an electrical signal to trigger a localised pneumatic system) are virtually instantaneous – milliseconds.
A dynamic brake would also take milliseconds to apply.

Is that used for both service and emergency braking?

The emergency brake still has an electric signal to the localised pneumatic system. The signal is just on/off though, not PWM. There is no variation in the control signal demand.
Purely pneumatic brakes would have a delay of about 2 seconds for brake pipe propagation.

I read a paper from Holland which claimed that application delays could be reduced by use of hydraulic braking systems?

That is true. They are very quick. The problem with hydraulics is that they leak. They are very risky if the fluid gets onto the wheel rail interface. There are a lot of concerns about hydraulics in railway applications because of that. It has been looked at, but it is difficult to prove that leakages would not affect adhesion performance.
The two advantages of hydraulics are that they are quicker to apply and that they are easier to control. If you have WSP, getting air out to release the brake is slow, but with hydraulics it is much quicker and easier. You can also work with much higher pressures and smaller equipment / tubing.

Qu 5. What is the difference between the current performance expected for 'service' or 'emergency' braking (both achievable rates and application delays)?
If you have to stop to a 9% adhesion rate, the brake manufacturer will build contingency into the brake system. On the track, you will therefore achieve a mean retardation rate of 9%. That will account for the application delays by having a higher peak retardation rate that counters the propagation delays and other losses in the system. The brake manufacturer will design to 10 or 11%. If the propagation delay was 2 seconds for example, at 225 that would mean about 120m. That would be compensated for in the braking rate later.
In the TIIS for West Coast Route Modernisation, it talks about 6% above 200kph and 9% below that speed. We will effectively achieve that. It will be delivered, or probably slightly better, even with propagation delays. If you just assume 6% and 9% for a control system braking algorithm, you do not need to consider the propagation delays on top of that – allowance for that will have already been included in the braking system. In practice, there will be some blending between the two brake rates, which means that higher brake rates will actually be achieved around the transition as well.
The only assumption in those figures is that there will be enough adhesion on the track to support the available brake effort.

Qu 6. What are the effects of adhesion?
If the adhesion level is only 5% and you make a 9% demand from your braking system, you will only get 5% braking. Railtrack do not define the adhesion levels for their railway. They specify braking performance in dry rail conditions. Someone needs to decide in the design of the train control system what adhesion can be assumed.
Leaves on the line and other rail contamination on the rail head will reduce the available adhesion. Railtrack have sandite trains and all sorts of things to try to keep the rail head clean.

In poor adhesion conditions (like leaves on the line) you may only have 2 or 3% adhesion available. That is usually only for short distances, but it still has a major effect. You need to build in some sort of contingency to allow for that. You can look at historical experience to see what contingencies are usually needed.

What potential is there to limit these effects / by doing what?

Sanders on the trains can help, but it has lots of downsides. One is interference with track circuits. Lots of sand on the railway also clogs up the ballast, reducing drainage and causing the track to deteriorate. Sanding also prevents lubricators (installed to reduce gauge corner cracking) from working properly by mixing with the grease. It is also questionable how effective they are at high speed. Sometimes they may increase adhesion by 3%, at others only by 1 %. You can not rely on them to make up your shortfall all of the time.

What about the use of track brakes?

Clearly Magnetic Track Brakes would improve braking performance in poor adhesion conditions. Their down side is that they are heavy and hard to maintain. I don't think that anyone would start with magnetic track brakes as a part of their railway system if they could help it.
Railtrack currently don't allow Magnetic track brakes on their system, presumably due to track works or raised check rails.
Is WSP provided on emergency and service braking, or just service braking?

It is provided on both.

Is the WSP fail safe?

The WSP attempts to maximise the adhesion of each wheel. If the wheel locked up, the adhesion from a sliding wheel is a lot less than one which is slowly moving. It is therefore safer to operate with WSP than without it.
WSP is trying to protect against wheel flats which are very damaging to both the train and the track.

Qu 7. What are the effect of jerk rates on passenger comfort and safety, including:
Jerk rates control how quickly you can allow the brake to come on. Most of industry use a value of $0.05m/s^3$ as the comfortable jerk rate value for braking and traction. In terms of safety, you could probably tolerate double that, $0.1m/s^3$.
You need to start by considering the quasi static braking forces. A steady 12%G deceleration rate alone would start to get uncomfortable.
Jerk rates that accelerate the body vertically and laterally due to track defects are well in excess of the values that I just quoted. The peak accelerations due to normal track irregularities are probably 10 time the values. However, those are the values classically specified today for traction and braking.

Is jerk rate limiting applied to emergency brake applications or just service brake?

The emergency brake would also be controlled so that the brake force is not applied at a higher jerk rate. The application rate is controlled through chokes in the pneumatics, so the control is still fail safe.

Qu 8. What is the current crashworthiness of rail vehicles?
It is probably best to describe crashworthiness in terms of speed. We have designed the new pendelino trains for Virgin to take a train to train collision of 60kph (either a 60kph train running into a stationary one, or two moving trains with a speed difference of 60kph). Up to that speed there will be controlled collapse of the end 1m of the body shells. At the front ends of the train we have designed to 3 times the group standard crashworthiness level and in the intermediate ends to 2 times.
The driving compartment would survive that crash intact (there is a collapse zone in front of it). The passenger compartments would also survive intact. However, the vestibules are designed to collapse as part of the energy absorption, so it would not be a safe zone.
In order to provide that level of protection for the driver, you need space, like the aerodynamic nose of our trains. Without that, it is much harder to protect the driver.

Do you think that the use of operation within relative braking distance could be justified on the grounds that any collision under worst case would not exceed the trains crashworthiness?

My personal view is that active crashworthiness should be the priority. Your signalling system keeping trains apart. Passive crashworthiness is there in case the active systems fail for any reason.
I think that you could put forward such an argument, but you would have to show that the risks are in the order of 1 in 1000 in a fleet life. Something in that order. It would be a very difficult job to justify it and get an acceptable safety case together.

How much does it cost to provide crashworthiness?

The engineering costs of crashworthiness are higher than if you do not engineer it. You have more calculation and validation work to do. If you look at a crashworthy body shell and a non-crashworthy body shell and considered the cost difference of building them, it would be very small. It is the fixed costs in designing the train that makes crashworthiness expensive. The actual unit costs would not be effected much. I don't think that you can really tie down what the cost is. It would vary from project to project. Maybe 0.5 to 0.75 million per project.
When you move on to interiors, like seats and tables, providing crashworthiness does start to cost a lot more on a unit cost basis. There would also be fixed costs for sled tests and validation and subsequent modifications on the interior crashworthiness. My view is that crashworthiness today is mainly about fixed costs, not unit costs.
You could also use seat belts to increase safety.

Qu 9. I have heard that the interface between ATP systems and train brakes is usually designed to SIL 1. Is this true?
No, train braking systems would be at least SIL3 or 4. All of the safety control circuits would be designed to SIL4.
I am not aware of any particular week points in the safety of brake system design. All single point failures would have to be designed out before a train could come into service.

Qu 10. There are certainly things that you could do to improve the braking performance that might not be very comfortable for passengers, but you could use them only in a real emergency in order to avoid a collision (such as use of high performance track brakes). What are your views on that approach?
The kind of issues you would have to consider are the catering trolleys flying off and hitting someone inside the train. Coffee spilling and scolding someone. You would have to prove that you would not injure people due to things like that in order to get higher braking rates approved.
Things like track brakes are allowed for street running, but not for normal railways. They provide a much higher brake rate and passengers do get thrown around. They can achieve something like 30% braking. They are vicious. However, that is a light urban transport system, not one that has people cooking hot food, hot drinks or any of the other hazards like that which exist on a main line train.
We also had a lot of problems on Manchester Metro link, We fitted a balise to make the track brake come on as trams left the Queen street depot. As a result of that, we had a lot of bogie component failures.

When you are trying to make the vehicle brake like a car, it is quite extreme. Could fewer or less powerful track brakes be fitted in order to increase train brake rates to, say, $2m/s^2$?

There is usually one per bogie, so I suppose that it would be possible. However, you would be increasing your axle weight. Magnetic track brakes are also dreadful from a maintenance point of view and they are not allowed at the moment on Railtrack infrastructure.

Qu 11.There may be things that you could do to increase braking rates that would not be fail safe, but if used in an emergency would actually be more effective at preventing a collision than the current emergency brake systems if they did work. Do you think that approach should be considered?
No. Potentially you could put more brake force on the wheels by using the dynamic brake during emergency braking, but you would still have to blend it with the friction brake in order to keep the total to 9%. Otherwise the adhesion would not be available to support the deceleration on the wheels. You would just end up over braking axles, risk wheel flats and still be limited to the deceleration supported by rail/wheel adhesion.
As things currently stand, you would also not get a train approved if its brake rate was higher than the standard.
Currently if a train has 36 axles, 12 of which are motored, you can get a third of the retardation rate off of dynamic braking. The rest would be supplemented by the friction brake.
You prefer to use the dynamic brake because friction braking causes a lot of wear and needs more maintenance. One of the major costs in maintaining high speed trains in friction brake maintenance. It is a life cycle cost issue. The dynamic brake is very forgiving. It is even better if you have regenerative braking available, you would use that. You can not always rely on regeneration though.
Most administrations have the same balance on what they believe can be achieved from the rail wheel interface. Braking is probably one of the most heavily researched areas for railways, yet people always end up sticking with the way it is currently done.

Donald Hayward BSc CEng MIRSE, MBCS (Head of Strategy, Alstom Transport Information Solutions).

Questions Raised by D. Woodland on 10/05/2001

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I started with LUL (for 17 years), working in the development of signalling control systems for the first 10 years. I started as a programmer, but then shifted more into understanding the systems, design of control systems and improving the design. This involved talking a lot to the operators to understand there requirements for the systems, etc. In parallel with that, I was working on safety cases for Westinghouse jointless track circuits that were used on the Bakerloo line (and subsequently the Central Line).

After that I was the 'Research Engineer', looking at innovative products that the Underground may have been able to use. Radar for checking train speeds, means of detecting when people fell between the cars of trains so that we could prevent the trains from moving, the development of service recording systems so that the performance of the service could be monitored and statistics automatically gathered and reported, etc. Latterly I was involved in the specification of a train identification and management system that involved tagging trains, recording where they were on the track and communication of the days plan to the driver. A general systems / operations role.

Since I have been at Alstom, I am head of department for Strategy – product introduction to the UK. That includes tools for the configuration of SSI (improvements to the automatic data generation), tools to assist the principles testing of SSI to automate the testing of code and report on anomalies. The automatic generation of control tables directly from a drawing that is a subset of the scheme plan. Also ICONIS, the new signalling control system, trying to capture data for that. Then there is the introduction of ICONIS itself, and in the future Smartlock to replace SSI.

Qu 2. The terms 'Signalling' and 'train control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system?

Signalling I use as a term closely aligned to the prevention of trains hitting each other and being derailed, more by conventional means. Colour light signals, points, track circuits and the control of the normal trackside elements that are involved in keeping the trains apart and ensuring there safe passage through junctions. That would include the interlocking. You can have signalling without any train control.

Train Control I would see as including signalling as a subset, but also the automation of train protection and operation as well. So a larger set of functionality that encompasses the actual control of the train to the constraints of the signalling. ATP, involving trainborne computer systems that monitor the trains movements against a braking curve or distance to go and can take controls away from the driver if he goes outside the envelope, I would regard as train control not signalling. Train control also takes in ATO, where the driver doesn't have to do anything, but is perhaps there to intervene if there is a problem. So, I would say that a train control system is a combined ATO and ATP that works within a framework provided by some form of signalling. It therefore takes outputs from the signalling.

I would not use the term Railway control system, but ICONIS could be regarded as a railway control system. It incorporates supervisory traffic management functions that manages the plan for the day. It also includes SCADA for control of power and telecommunications (signal telephones and radio) and control of auxiliary station equipment (escalators, etc.). An over reaching system of control that provides input to the signalling at a lower level.

Qu.3 Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

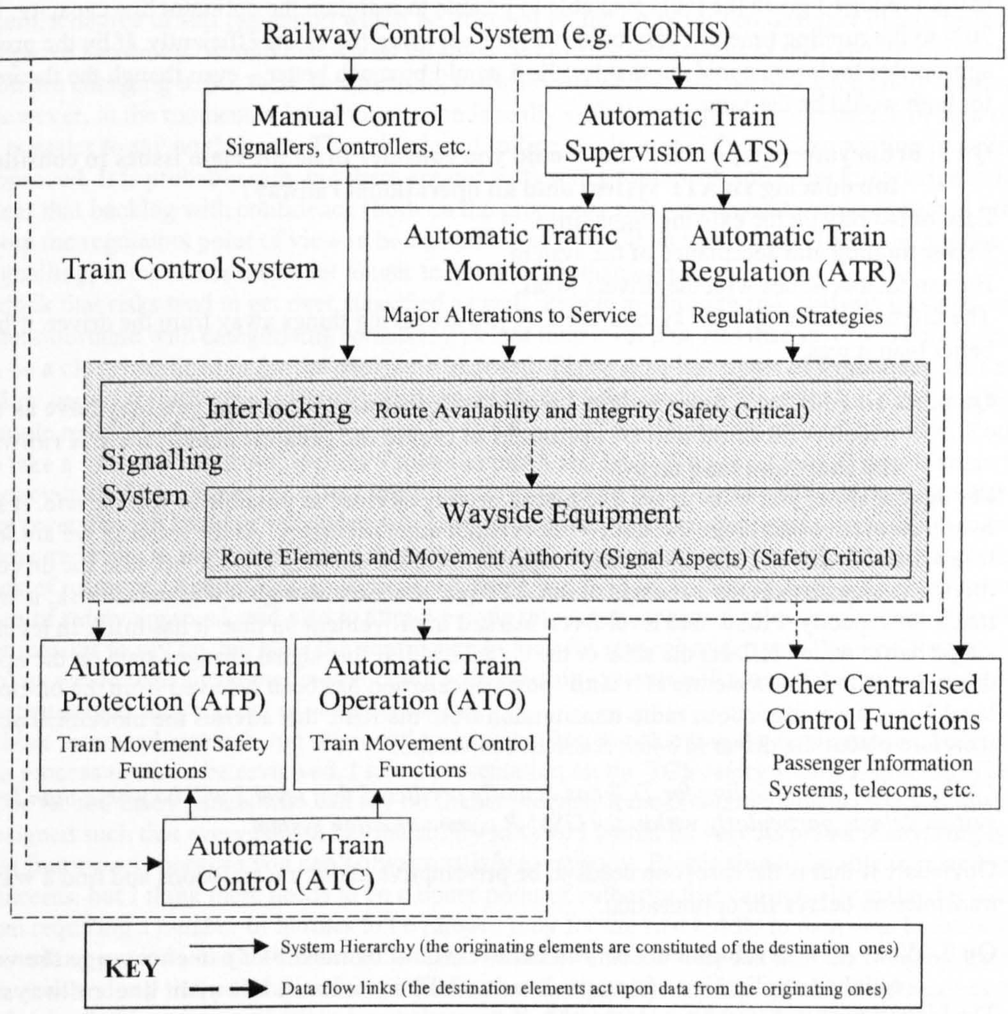
ATS is a system to automatically manage the service to a timetable framework, giving the operator the ability to adjust as events require. The timetable starts off at the beginning of the day as the plan, but pretty soon you will deviate from that. ATS will automatically route trains to their required paths, drawing problems to the attention of the operator (if there are service gaps or things are not moving), but giving the operator the chance to intervene and change the plan.

ATR is a much smaller subset of ATS used for evening out intervals or holding trains to their departure times. It is more about evening out gaps.

If the order or routing of trains has to be changed, they have to be turned short or have their journeys extended, that would be ATS. It is outside of the scope of ATR.

There is no direct link between ATC and ATS. ATS issues requests to the signalling elements of train control to ask trains to go or not, and to get routes set up. Signalling control provides a safety layer between those requests and the ATC system, which will actually allow the train to move.

Could you draw a diagram to represent the hierarchies in control systems as you see them?



Qu 4. How would you define the capacity of a Railway system (and what units would you assign to your definition)?

There is a hierarchy of metrics that you could use. The best one I have seen is to do with passenger journey capacity. The idea is passenger throughput – people per hour – against certain journeys. You can not just measure the number of people past a given point, because it could be a very wide pipe with a slow flow. You need some sort of measure of journey time too. So, people per hour in a given direction but also a measure of journey time or speed in combination with that.

There is a metric that LUL use for working out whether you have made anything better or not 'notional accumulated customer hours'.

Within that, there are more specific measures such as signal headway (the time interval between successive trains), but if you have trains that give you larger numbers of people and therefore give you better capacity for a given headway, and if the trains are travelling faster, it also gives you a better capacity. So Headway is a rather narrowly focused measure.

Something about stability needs to be included. Something may be theoretically achievable, but once you have nudged it from that theoretical optimum (if it goes unstable and becomes extremely un-optimum) then the capacity measure is totally optimistic. It needs a resilience factor to allow for a margin of perturbation from which you can still recover. I don't like building recovery margin into timetables as we do at the moment. That is just building in low capacity all of the time in order to be able to cope with problems when they do happen, which is not all of the time. I would rather that better tools were given to regulate the service and manage perturbation than under use the assets. Regulation is a method of achieving stability. Given that in the past a Signallers only means of regulating trains was to hold or cancel them, fairly basic low level things, you couldn't really have a recovery strategy apart from building in a margin to the timetable. I think things have now got a lot better than that. With the ability to now predict what is going to happen and introduce systems that are able to suggest strategies for overcoming a problem, we can start to eat away at the margin that is normally included within the timetable to allow for recovery and actually squeeze more service through. You could improve the stability of the same optimum line capacity through the provision of more robust support tools for manipulating and managing the service during perturbation. You could measure that as the recovery margin required given the tools available to be able to maintain the optimum line capacity. If you just add 20% to the running time of every train, you are not using the assets efficiently. If by the provision of appropriate tools you could cut that to 5%, it would be much better – even though the theoretical headway for both would be the same.

Qu 5. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

Ease of overlay on the existing signalling.
Driver training and acceptance of the system.
Human factors issues with the drivers MMI.
The safety case for the ATP system itself. If you are taking things away from the driver, it has to become better than it was.

Qu 6. As an additional point to this, I would welcome any views that you may have as to ways in which ATP could best be optimised to create the greatest benefit for UK railways in both the short and long term?

For optimisation, you want to get it to permit driving as close as possible to what is safe. It must not be overly restrictive and brake the train when it is not quite necessary. At the moment we are looking at TCS levels 1 and 2. Level 1 seems to reduce capacity from the existing system, because the driver is better at driving to signals than the envelope of the ATP permits. So although it will reduce risk, it will not do much for capacity. I think that level 2 is a marked improvement on that. It has infill. In level 1 you pass a single balise which reflects the state of the signal head. If that signal changes state on the approach, the driver is not free to accelerate at it until some cancellation has been received from the on-board system. In level 2 you have continuous radio transmission from the RBC that advises the movement authority and therefore places the driver in a less restrictive envelope.

I believe that the modelling for TCS has actually predicted that level 2 will be worse than level 1 due to system delays, particularly within the GSM-R communication system.

Obviously if that is the case, one needs to be pre-emptive about transmissions and find a way to attack the transmission delays for optimisation.

Qu 7. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

The biggest problem is getting acceptance. If the system got national acceptance and was available for use, then the questions of cost and time / resources to apply it. So it needs to be easy to fit – to a variety of rolling stock. It is not much use if you need a large space in the cab of a train which is perfectly air conditioned, so that it can not actually be fitted to a variety of stock. At the moment ATP systems tend to be introduced with new rolling stock, so it is relatively easy to fit. We must have many different types of stock in service throughout the network, and if we have to wait for the replacement cycle of that stock to come around in order to fit ATP, that would be a massive brake on the process. So the interface needs to be easily adaptable for fitting in different types of stock and the environmental requirements must make it suitable for fitting in all types of stock.
Trackside, there are issues with the GSM aerials springing up everywhere. There is an awfully long planning processes to go through to be able to get sufficient coverage. So, from an environmental point of view, it may be better to look for some other kind of track to train communication than GSM-R. We probably also need to go straight for level 3, rather than level 2 on the entire railway as it will be much easier to fit. The interface to the signalling is much less invasive - the train reporting its position back to the RBC rather than having to interface with the existing interlocking. Just as there are many different types of rolling stock, there are also hundreds of different variants of signalling – all of which would require a specific interface in order to be able to capture the information for level 2. So level 2 on the national network would be more difficult than TPWS – and that has been difficult enough! We really need to try to minimise interfaces to the existing signalling.
It might be useful to retain some form of conventional signalling to act as a fall back mode and allow mixed running of fitted and un-fitted stock during introduction. Quite a challenge I think, particularly if we don't go for the higher level – level 3.

Qu 8. As an additional point to this, how do you think the trade off between safety and operational flexibility can best be managed?

The commercial push will come from operational flexibility – improving throughput.
There must be independent regulation to ensure that safety is not compromised for commercial reasons.

At the moment, safety approval panels and the HMRI act as regulatory authorities to stop changes that would introduce new risks. There are also the yellow book safety management processes. By following them, a degree of self regulation would be imposed on the proposer and designer of any measure to improve flexibility. Within that would be some sort of risk assessment of the existing situation and how you are changing it. So, there is something there already that should work if adequately resourced. However, at the moment, a lot of innovation is really stifled through lack of resource or lack of courage. It is easier to say no than yes. There is a lot of rolling stock around the country waiting in sidings to get approved. It is probably safe, but there are not sufficient people with sufficient knowledge to be able to clear that backlog with confidence (both on the proposers side – to understand the approval process – and from the regulators point of view in being sufficiently confident to say yes). The same applies to signalling, some issues take a lot longer to get through than others.

I think that risks tend to get over classified as well. People go through risk analysis sessions, and are very uncomfortable with categorising something as less than level 2 or less than level 4. There does not seem to be a clear justification for the decisions imposed. There is a plan, or set of techniques, that are meant to allow you to identify what safety integrity level should apply to each function, but there is a tendency for people to decide that if a system has one level 4 function, the whole thing must be level 4. You then have to take a lot more measures, it costs 5 times as much and takes 5 times as long. There is an awful lot of money in that for a lot of the safety industry, so you end up being better off sticking with what you have got (even though sticking with what you have got may actually be a lot worse in reality than taking measured risks to move forward). I think that the balance currently chokes off innovation either due to lack of resource or people being unprepared to say yes. So I think that we need a bit more resource in the area of safety approval, and also to attract people into safety approval roles who are not always looking for reasons to say no, but have the confidence to look for reasons to say yes. We won't move forward unless we do that, and sometimes sticking with existing systems will actually be giving us greater risk – particularly as they run out of supportable life.

I think increased research and improved training are key to achieving the requirements. I also think that the process needs to be reviewed. I saw a presentation on the TCS safety approval process. There seemed to be far too many panels who can say no in that process. It looks totally bureaucratic and always designed such that every person has the ability to veto. I would be very surprised if anything gets through that framework, because you can't always satisfy everybody. People should be able to raise legitimate concerns, but I think there needs to be a better point of authority that can actually make decisions rather than requiring a number of hurdles to be jumped only for the first hurdle to then veto because of something put in by the fourth hurdle! I think we need someone who can take the thing by the scruff of the neck and say that people can have input but not necessarily veto input. I think there needs to be one authority that has veto, and others can only suggest. The ultimate authority must then have confidence about what they are saying and consider all of the views. I don't think that you can allow multiple committees to have veto and still expect a system to pass through them all in a reasonable time scale.

Qu 9. What information do you think:

A. Is required by a driver in order to perform the tasks asked of him

Departing a station: Doors have to be shut with no one trapped or in risk from them. The driver therefore needs some sort of confirmation that the train is ready to depart, and a means of communicating that he is ready to depart. At the moment, he has to look at his starting signal, maybe gets a bell from the guard and maybe sees someone waving something on the platform or an RA. In some places he has to press a ready to depart plunger outside of the cab. Basically, system to carry out these functions are pretty important. Obviously, systems to do this would have safety implications. If we are saying to him that it is safe to go and it is not, but he goes anyway, we are in trouble.

Between stations: Indication of what route is expected to be taken (in case it is the wrong route / one that he does not have route knowledge for). The permissible speed for the route– perhaps as a speed profile, and his current speed / location against that. Alternatively , a graphical speed profile that includes details of where his train is against the profile and his current speed (so that he can see how he is driving compared with the limit) would remove the reasons for telling him which way he is going.

Informing Passengers: Information about the reasons for delays that may be experienced, proposed revisions to the journey – in time or route. Any important connection information that passengers might need to be advised of. These things should not be flashed up in front of him, but available on request.

First Line Repairs: I don't think that everything about the train should be flagged up to the driver. Only those things that he is really expected to be able to manage and either repair or use a fall-back method to get around. The driver is going to be the first line maintainer of the train, so he needs diagnostics to an appropriate level.

Communication with central service control: The means to effectively report any incidents or query any instructions that he has received.

B. Would be beneficial to a driver to assist him in the performance of his tasks

Between stations: It would be useful if the system could automatically detect, using the drivers log in data ,whether the route given was the correct one / one for which the driver was qualified. Station stops should be reflected in the speed profile, so that even though the signalling is happy for him to continue, the profile tells him to stop.

Informing Passengers: Automation of messages, rather than requiring the driver to worry about it, would be very useful. When the train is standing at a signal, it may be more appropriate to let the driver handle announcements, but when moving it would be good if announcements occurred automatically – to avoid distracting him.

First Line Repair / Emergencies: Most of the job is routine, maybe once a year a serious incident will come up. If there are scenarios detectable by the system, a bit of prompting about what to do would be useful. I.e. if a passenger alarm is pulled in the third carriage, the MMI could prompt not just that it has happened, but a list of actions to take as a result.

C. Would be detrimental to a drivers ability to perform the tasks required of him

Departing a station: expecting him to look at video pictures of the rear of the train whilst departing. He should be looking for hazards ahead. It is also a bit unreasonable to expect him to look at multiple mirrors or monitors.

First Line Repair: Not all fault finding / diagnostic information. It would just clutter the information in the cab & encourage him to do things he is not trained to do. More information should be available for trained maintainers.

Qu 10. How do you think that your answers to question 9 would change if the system included ATO?

I don't think that it would change much, since he is supervising the ATO, he ought to see similar displays. I think that you should give the same display regardless of whether he or the ATO are manually driving. That way he can watch how the ATO drives against the profile and learn the pattern of how it drives. That would then help him to improve his driving style when he has to drive manually from time to time.

Interview by e-mail with:

Andy Heath BEng (Infraco BCV).

Questions Raised by D. Woodland on 17/04/2002. Responses given by S Rodgers on 22/04/2002.

Qu 1. I am interested in the way that speed limits and restrictions are set on both UK main line and metro railways. Do you know haw this is done on LUL?

I can give you a few snippets of info on speed restrictions on the LUL network, from my signalling viewpoint:

- Everything is based on Vmax, which I understand is a maximum speed for passenger comfort, based on track geometry, the values being derived from some practical tests on the main line (back in the 50's I was once told).
- On the LUL manually driven lines, the PSR's can be assigned right up to the Vmax value (should it happen to be an exact multiple of 5mph), and will extend as far as the longest passenger train's length for that line beyond the restricting feature. PSR's will only be signed where the fastest train starting from 35km/h at the previous station can actually reach the PSR speed.

Qu 2. I assume that the calculation method must include some allowance for over speeding through a section. Is this the case, and if so, what is the allowance?

A degree of non-compliance with the signed PSR is allowed for in overlap calculations: the overlap curve graph we work from assumes +15km/h (if I remember rightly; I don't have a copy of the spec at Telstar unfortunately).

On the Central line, the important parameter seems to have been [Vmax + 15%]. The simulations against which the signalling was laid out show a train travelling at 'headway' speed from rest at the previous station; the ATP blocks were then positioned and max.-safe-speed codes allocated such that even if an ATO controller fails in full motor mode the train will be tripped and speed brought down such that the [Vmax + 15%] profile is never quite infringed.

My Vic line notes must be at 30TSC as well.

If you need any more info from a signalling point of view, let me know and perhaps we can fix a meeting with someone more knowledgeable than me. From Track, the only person I've worked with on such things in recent years is Peter Kinselley, who's a helpful sort of chap. He's has ended up in SSL.

Barry Hills (Operations Principles Consultant, Long Melford Associates).

Questions Raised by D. Woodland on 08/03/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I started at 16 years old, as a Second man – the assistant river. The scheme was then that you worked as the
Second man until you were old enough to take the exams to become a driver, which I did at 21 years old. I then became a train driver at Stratford in 1976. In about 1985 I began to feel that train driving held limited opportunities, so I became an acting supervisor (relief man looking after men and machines at Liverpool street), whilst still driving as my main job. Then after a couple of years of doing that. I got a permanent appointment at Shoeburyness Depot, where I was movements supervisor (train crew, sidings, signal box and station) for about 15 months.

Then I went up to London as Controller on the LT&S railway. This was dealing with the co-ordination of train movements, making decisions on which trains should be cancelled, held or turned back when disruptions occurred to the service.

Then I moved into my first management post, responsible for freight performance. Then Operations manager for freight, where I had around 350 staff in East Anglia.

After that I was promoted back to London as HQ train crew manager for a freight company, dealing with resources.

I worked on various other projects, including the development of the Class 92 Loco operating manual, and ended my time working for the railway as Operations Standards Manager, with my own operations support group, supporting day to day freight operations.

In July 1988 I set up my own company. I then contracted to Railtrack for 12 weeks covering work in their standards and safety department.

That was followed by various other roles outside of Railways.

I then returned to Railways for Railtrack's year 2000 project until January 2000.

Since then I have been working on the West Coast Main Line TCS project, within the Operational and Signalling Principles team.

Qu 2. The terms 'Signalling' and 'train control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control system'?

Signalling to me is the controlling of trains by fixed trackside signals, giving the driver an indication of the point that he is allowed to go to. This could include where you don't have a block system, but instead give the driver a reminder that he does not have certain forms of protection, so things like 'low speed' and 'as far as the line is clear' come into play. To me the Signalling system is really almost a mechanical interface between allowing the traffic to run from point to point and routing it correctly.

Train control I believe is a bit more sophisticated than that. In the TCS case, it is an overlay on the existing signalling system, but train control can do many other things. It could for instance make trains travel at different speeds to alleviate congestion. It would certainly guarantee that certain safety parameters won't be breached, so while drivers are human and make errors, the train control system would prevent any major consequence arising from the Driver's error. So if the driver makes an error, we can stop that turning into a major event, which is a level of assurance that we do not have under signalling, because we rely on the drivers judgement and rely on communication. Train control is very much a system that you can not get around, which is why I prefer it.

The difference for me is that one is almost optional. The trains movement is only controlled by a drivers interpretation of signals, sometimes wrongly, or the wrong routes being set. There is some leeway for human mistakes in almost every part of the system. It is about peoples' knowledge of routes and their judgement.

With the train control system things are much more rigid. People will get used to the approach speed curve to a particular signal, and they are usually very adaptable to learn that, but Train Control simply won't let you step outside of it.

There is very little point in asking someone to operate a switch, when it could be done automatically by the control system., and the person can be there as a supervisor, ready to step in if there are problems and the system is degraded. SI I do not see any problem with a fairly automated system where the train control system does actually control the trains, not just from a pathing point of view, but applying power and braking to trains too. You get absolute optimums from that, where as a human can not perform optimally, because people react differently. The system could deduce braking characteristics and adjust performance accordingly. Drivers do try to do that, but can not do it as effectively. To me, even simple things like cruise control help a driver to follow a target speed much more accurately that a human driver could. Railway control system would perhaps include automatic features, such as ATO.

Qu 3. Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

Regulation is reaction to what is happening to keep the service running. So within the elements of the system you may have to alter the train stopping patterns, make the trains longer, change crew schedules. Basically regulating the perturbed railway. The system only has a certain capacity, and really we should be able to determine all the time what that capacity is, and have almost ready made solutions for adjusting it. It has to be fairly high level, because the overall picture needs to be seen with as much knowledge as possible in order to make the best decisions. An automated system ought to be able to do that.
Supervision I understand in terms of what we are trying to do with TCS. Supervision and Regulation conjoin. Supervision uses the train control system to act at a lower level than regulation, to keep the train service operating as it should. Avoiding holding trains just out of station areas, and that sort of thing.

Qu 4. How would you define the capacity of a Railway system? Please state what you would consider to be appropriate units for your defined measure.

How many trains you can run with a certain degree of reliability within a given time. So, if you want to run 20 trains at 100 mph, the same railway may be able to run 40 trains at 50mph. So the capacity of the system is something like trains per hour in a given section. Other things would be movements stop the railway working: junctions, crossovers, etc. which take back the performance that you could get running on plain line.

Qu 5. Given the low costs of TPWS as compared with a comprehensive ATP system, do you think that TPWS can be seen as achieving the optimal ATP solution in terms of cost / benefit, or are there other possibilities? Please explain why you hold the views that you do.

I have read a lot of papers on TPWS, and something does not seem quite right to me. Firstly, it will not be put in everywhere, so when a driver makes an error and passes a red light on plain line due to lack of attention, it will not prevent that, because it would not perceive a signal on plain line as a serious risk. That is its Achilles heel. I can think of a number of past accidents that simply wouldn't have been prevented by TPWS.
The other point about TPWS is that it relies on equipment and cable in the 4 foot. Vandals and maintenance activities will cause failures there. The more you put physically in locations on or about the track, it causes a problem. Anything that prevents or mitigates SPADs is great, but I think that TPWS is a political answer. It is a half way system to the full ATP and is now an expedient to mitigate the fact that full ATP was not fitted throughout the network, as recommended by the Hidden report.
I am not sure about details of its performance.

Qu 6. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

I have experience of the cab signalling system TVM 430, which is not particularly user friendly, but is a particular solution that suits high speed lines (such as the Channel Tunnel). In my view, we have to keep it as simple and understandable as possible. We have to make sure that people are not given choices within it. So where we are talking about selecting modes, I think that should be a system issue. The driver may need to confirm acceptance, but I would not give people choices, just so that they can make the wrong choice (even if that only results in performance issues – the train not working or moving). I am sometimes confused by all of the different railway 'modes'. Shunting, reversing, calling on, subsidiary,... The railway is very simple. You either have a piece of track where the signaller has allowed you to go and controls that piece of track, or you are allowed onto the piece of track on the special condition that you can only go as far as the line is clear at low speed. All of the different terms used just confuse that.
Today's drivers are more and more lads who will do 9 to 12 months training and then become the driver of a train. We need to give them the main stream options every time so that they understand it. If they don't have the protection of ATP for any reason, they will have to rely on skills that they may loose due to built up reliance on ATP. I think that was what the report from Sweden was saying *(Kecklund et al. 'The TRAIN-project: Railway Safety and the Train Driver Information Environment and Work Situation – a Summary of the Main Results', Signalling Safety Conference, IQPC, 2001).* This could be an issue, which may mean that we have to ask drivers to demonstrate from time to time that they can drive a train normally on a simulator. So for me, driver training is very important. Simple to understand the system, what it will do and their role in it. Otherwise we will confuse the issue.
I would also like to see a consistent solution across the whole railway.

Qu 7. As an additional point to this, I would welcome any views that you may have as to ways in which ATP could best be optimised to create the greatest benefit for UK railways in both the short and long term?

If we get it right the first time, then we can build on that. We are already introducing thing like enhanced speeds.

In my opinion, to be optimised would require a radio based system, because that is the absolute degree of train control. The decision to stop a train in a particular area can then be taken by either the signalman or the driver, which to me would be a great added factor. We still don't have a decent radio system at the moment, and rely on people stopping at telephones.

Consistency is another main point. We currently have about 8 different signalling systems in the UK. I think that we have to get ATP right for it to remain in place as a single best solution for all time. That might be difficult,, so what we are looking for is to go right to the edge of technology and work out what we can reliably deliver.

Finally, reliability. If we don't deliver a system with high reliability, people will start to distrust it and to cut corners to get around it. Isolating systems and stuff. It then becomes sub-optimal.

Qu 8. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

I think that wider / quicker application is more to do with politics and the will to spend more money than developments. We need a rolling programme for ATP fitment across the whole network.

As for developments, today the railway is not very different to the way it was 150 years ago. We would do much better, more reliably, with radio based systems rather than fixed ones. To me that would be a much better way of doing things.

We need to look at getting the costs down.

We also need simple solutions. Perhaps a version of TCS that can be simply applied to the simple parts of the railway, preferable getting rid of lineside signals and having a fallback mode of reflectorised marker boards with block to block working.

Qu 9. As an additional point to this, how do you think the trade off between safety and operational flexibility can best be managed?

I think that operational flexibility may have to suffer a bit to bring in improved safety, but there will also be a need for the existing railway regime to move to ward the new world, in the sense that some of the railway could be better managed. I mentioned the difference between shunt routes and calling on routes. The only reason that the fixed equipment for those is different is that we don't pass a red aspect – even a miniature one on a shunt signal. To me we should be talking about shunt all the time. Calling on is really a shunt to an occupied section of track. The Railway may have to change its way to move towards simplifying what it does. It wasn't until I started working on TCS that I came to appreciate this. Until then, I just accepted things as being the way that they are. The Rulebook in particular probably needs consideration.

The trade off in terms of operational flexibility in stopping a train at a particular point will change with the tighter regime of ATP, but I don't think there is a big issue there. Drivers will learn that and get used to it. There are bound to be moans, but Drivers will recognise that this is really saving every ones neck if they make a major mistake. Flexibility will suffer ever so slightly, but not so much as everyone seems to think, because it isn't black and white. All of the factors that make a driver drive a train to a certain speed at certain locations differ with each train and driver. Therefore, you only ever get an average, and the operational flexibility that we may lose due to ATP margins will probably be well within the normal range, rather than outside it. Perhaps modelling would give a different view on that.

There is a view that some of the safety requirements on railways (based on developments after years of accidents) don't really contribute to safety any more, but do impact on operational flexibility. Do you think that it would be possible to make changes to railway safety on this basis?

You are right that we have had a reactive railway over the last century. Someone examined an incident and looked to see what we could change in terms of management systems, etc. That resulted in a safe railway, but one that does not look at the overall system. We need to think about what we can now control better, or in a more defined manner. I can think of 1 or 2 examples – such as open level crossings (where we used to show a flashing white light to show the crossing was operating normally. We now show a flashing red light, that changes to white when it is safe to proceed. Before, if the crossing failed, the light just went out and the driver was supposed then to stop). Really we should be getting rid of crossings. People and trains should not be crossing at any point unless physically separated. The Railway can be quite regimented, but the roads can not.

If you didn't increase risk and could demonstrate that, maybe because the risk is poorly managed now, the change would be acceptable. After the accidents in the last few years people won't accept anything other than ultimate safety, and they are prepared to pay in terms of journey times or anything else. As an example away from train control, people were falling out of trains, so they put central door locking on trains. Everyone on the train now has to wait for a few seconds when the train stops for the door to unlock. People do accept that as an improvement to safety.

I think that you need to consider the overall railway system, and the signalling as a part of it. I don't think that we trade off at all on the Safety side. Obviously, if you could demonstrate that your changes would not increase casualties that might be acceptable. You could then change the signalling as long as you could demonstrate that the railway ended up with a degree of safety as good as today.

Another example, would be moving block, or even allowing trains to follow each other at closer than braking distance. After all, we do it today with goods trains. Why is the driver's life in a goods train any different to the driver and people in a passenger train? Being pretty simplistic, we do it in our cars every day, so why not on a railway train? We think that the risk is perfectly acceptable in our cars, or we wouldn't get in them to come to work. We also rely on other people in that. The roads are completely stuffed in this country at the moment, so we have to pay to develop our railways in order to supply the service needed by the social need. I don't think that anyone costs my 2 hours each way to get to work by car. I feel that we must look at the wider cost benefits in society. Watford as an example, is crying out for a better public transport system. The roads are crowded, and many people would go for a viable alternative – reasonably priced, safe, reliable and clean. The railways were getting to a point where they were starting to do this before privatisation. People also expect safety of course, but the railways are far safer than the roads. There is a social side to this. If we could take journeys off the road we would decrease the risk, even if the railways risks increased.

Qu 10. What information do you think:

A. Is required by a driver in order to perform the tasks asked of him

He needs to know the speed that is required, the trains performance in acceleration and braking, how far he can move safely, and in what manner (whether an absolute authority or a conditional one). There are some add-ons in today's railway of course, such as AWS and trackside signs, but the driver knows those from his route knowledge. I am not sure signs really make the railway safer. They are just reference points. The driver does need them, but they could be just a white post beside the track.

B. Would be beneficial to a driver to assist him in the performance of his tasks

I can see the benefits of having ,if you like, a cab repeater system to show signal aspects on today's railway, if we were not going to ATP, because it would fix is the drivers mind what is coming along. There was a trial on the Southern region that had a set of miniature lights in the cab.

For the limit of authority, a set of trackside markers and a braking profile in the cab to support that would be beneficial. I also would like to see the information delivered differently. It is either visual or a system of whistles and bells at the moment. I think that other audible information would be better, like the way the aircraft industry due it. Consistent instructions spoken to the driver by a 'nagging nora', saying 'do this'. Very clear and unequivical, just give the delivery of a particular message. We could use those sorts of things with ATP, rather than trying to translate a graphic representation through the eye to the brain, we could actually tell someone what to do. With the TCS MMI, we are starting to get to the resolution on the screen where we are asking the driver to use it like a computer game. Personally, I have to stare at computer games intently and get tired eyes very quickly. The driver should be asked to react to spatial stimuli –where they are and the feel of the train, not a computer screen. An audible instruction backed up with a visual display would be much better.

The driver will actually be fixing the point in his head, and knowing where he has to stop. Kilometric markers every half Km along the railway would help the driver to quite happily move his train over 2 Km from 125mph to zero.

C. Would be detrimental to a drivers ability to perform the tasks required of him

A poor display. I like to see what other railways do on displays. I think a poor display and getting the driver to look at it for more than a cursory glance. I don't mind a back up, like we use AWS now, which acts as a reminder that you can almost subliminally look at as a reminder.

I would not like to see loads of noises in the cab. I don't mind voice commands or advice, but I don't particularly like the whistles and bells approach. It is quite confusing, and if you are going to give a message, you may as well give the message – we have the technology now to do that.

Qu 11. How do you think that your answers to question 10 would change if the system included ATO?

I think that the information sources would then be for comfort, but with ATO could be a bit more interactive with the driver. Non repetitive, reactive to the changing conditions on the railway, but keeping the driver interested and a part of the system. He needs to feel that he is not just sitting there reading his

newspaper, but actually has a job to do in terms of train monitoring, monitoring passengers, knowledge of system degradation on his train or around it. The visual indications could be as simple as a sign for clear and distance to go, or not clear. All the rest of the information could be suppressed under ATO.
The driver needs to be able to intervene quite comfortably, and to keep his own level of skill up.

Supplementary questions raised by e-mail on 30/01/2003. Answer received 10/03/2003:

Qu 12. I have got myself a little confused about how a driver is supposed to react to flashing aspects. I know that the idea of FYY is to allow a higher speed approach than would be possible with approach release from red, where the conditions required for approach release from Y can not be satisfied. This is done by letting the driver know that the slower speed diverging route is set ahead of the train. I would not, therefore, expect the driver to have to react in any way on passing a flashing double yellow aspect (the advanced indication). Am I right in this assumption?

For each driver the reaction or "mental note" of the flashing aspect may mean something different - but in reality the flashing aspect is a confirmation of position and following action. The physical action may or may not be initiated at the sighting of YY(F), but a driver will consider a YY (F) in exactly the same manner as an advanced warning of permanent speed restriction (some are signed - some are not). So (in this example) if the reduction in speed was not signed (due to the overall percentage reduction being below a value that I forget) -the driver would still have a mental note of where he is required to apply braking. In simple terms all stimuli are input to the central processing system, which selects the appropriate course of action based on inputs received on the day, archive material, experience etc!
I have spoken to both Peter Kirk (my Chief Traction Inspector) and Nigel Hutchison (a driver) they agreed in principle with my opinions

Paul Hosey (Rail Operations Specialist, St Enoch Partnership Ltd).

Questions Raised by D. Woodland on 17/04/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?
My railway career began as a Driver's Assistant in October 1973. After 2.5 years I progressed to Driver, which I did for 7 years. I then moved to Train Crew Instructor, Train Crew Inspector and then Train Crew Depot Manager (local management of drivers, drivers assistants and guards – recruitment, training, staffing, local discipline and negotiation with union representatives, accident investigation work, management of emergencies, etc.) by 1983. I was Train Crew Depot Manager at Ayre, Manchester Sheffield and Brighton. I then transferred to operations as Area Operations Manager at Brighton (looking after station staff and signalling staff), covering the south coast and Sussex areas of the south central division. Then after 3 years I moved to safety audit work for intercity at York. I was then responsible for introducing the ISRS safety audit system to BR.
I left BR in 1994, and have been doing railway operations and safety consultancy work since then. Rules and regulations development, safety case development, general safety management and project work with railway contractors, TOCs, Railtrack, LUL, Eurotunnel and now the West Coast Main Line TCS project as Operational Principles Team Leader.

Qu 2. The terms 'Signalling' and 'train control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system?
Signalling is the process of routing trains to where they have to go, whilst keeping them separated.
Train control is to do with supervision of the actions of a driver. It would be superimposed on the signalling system
Railway control is not a term I have come across, but I would understand it to describe a system for controlling the overall running of traffic. So it would include signalling, train control and regulation.

Qu.3 Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?
Regulation has a very clear and distinct meaning. It is the management of train movements to minimise delay to the whole service. So you may in fact delay a train, or allow a slower train to proceed a faster one, if the overall result is minimum delay to all trains.

Supervision has a direct relationship to what happens in a driving cab or the signalling centre. It supervises a human is doing and intervenes if he does something that he shouldn't.
There is no connection between regulation and supervision.

Qu 4. How would you define the capacity of a Railway system (and what units would you give your definition)?

I am accustomed to defining capacity in terms of train movements per hour, or minutes between each train, over a given stretch of infrastructure. That would be somewhat related to the speed of the trains and the spacing between signals.
You could also consider capacity as tonne miles, but as an operator I prefer trains per hour.

The problem with tph is that it is point specific, so how can you measure capacity across the network?

The capacity is always governed by the speed that you can make trains follow each other, so with a junction say, that has a 100mph approach from one direction and a slower approach from the other direction that causes delays, the capacity becomes the lowest common denominator. You have to allow for acceleration away from the slow speed approach.

Qu 5. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

The first consideration is can you still run the traffic that the railway needs to run. A train control system is unlikely to not have any effect on throughput. The systems that I am familiar with would all slow trains down. If they slow then down too much, can the railway operate the service required with that constraint? That is really the main issue in my view. The effect on throughput of traffic.
Secondly, will the driver have to constantly fight the system in order to run the train as close to its maximum potential speed as possible, or will it guide the driver. There is an argument that drivers in the UK who are experienced in driving would end up continually fighting the system, because they know that they could drive faster and still stop safely, but the system has to assume a lowest common denominator and will not let them. New drivers accustomed to ATP might be more likely to use it as a guide, but then if ATP goes down they would be lost without it, not knowing when they have to brake without it. Provision of ATP could make experienced drivers from the old system frustrated by its constraints, whilst newer drivers become dependent on it.
Thirdly, unless the system is incredibly sophisticated, it will have to calculate approach speeds, deceleration rates and stopping points in accordance with the worst case common denominator for weights, brake forces, rail conditions, etc. I do not believe that a system could make all of the calculations accurately enough whilst still always failing safe so as to meet safety requirements. That is another issue to look at carefully.

Qu 6. As an additional point to this, I would welcome any views that you may have as to ways in which ATP could best be optimised to create the greatest benefit for UK railways in both the short and long term?

Ultimately the first issue with ATP is prevention of SPADs, especially the higher risk ones that could result in collisions. However:

- The vast majority of SPADs are minor misjudgement and only a very small proportion are a complete disregard.
- ATP will eliminate accidents like Southall and Ladbroke Grove, but not other types of accident. For instance, a shunting irregularity may cause the main line to be fouled by a derailed vehicle, there are track or infrastructure failures. These sorts of things will not be protected against.
- If you look over the last few years at the number of fatalities caused by SPADs against those caused by other issues, to me it is questionable whether ATP is the best way to spend the money. It may be that we should spend more money on preventing bridge strikes and vandalism.

ATP is a step towards providing the driver with the same degree of automation that the signaller has had for many years. It is therefore an inevitable move, but at the moment its installation is politically motivated. Safety is the main purpose of ATP, but it is not really the best measure for safety.
I have yet to see a system that enhances operability and capacity of the railway. To optimise the use of any ATP system, I think we would therefore be looking to enhance capacity. There is the potential for huge improvements for reliability is you could reduce trackside equipment, say with a TCS level 3 type system. That would effect capacity to some extent, as you would not have to allow so much contingency for recovery time in the timetable. There would therefore be savings in installation costs, maintenance and failure rates, but it still wouldn't really enhance capacity. It may allow increased speeds – but increased

speeds just reduce capacity as you have to allow longer distances between trains. So, I think that a system that could improve capacity would offer greatest benefit.

Do you think that TPWS could be seen to be an optimised ATP system in terms of cost / benefit?

TPWS has the same disadvantages as ATP. It is a stopgap measure that has maximum benefit for minimum cost. It may offer extremely good value for money, but because it is not fail safe, it has a deadly disadvantage. It will not supplant ATP because of that, but it is a good temporary measure until ATP becomes available nation wide. It also significantly effects train speed, slowing everything down to the lowest common denominator. It has two settings, one for freight and the other for passenger. The problem with that is that a loco hauled express passenger train has a braking curve completely different to a modern multiple unit, but everything has to be calculated on the basis of the same assumed braking rate. It would be nice to have 4 or 5 different levels, but at the moment there are only passenger and freight. TPWS does not improve capacity, it limits it.

Qu 7. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

ATP systems tend to operate rigidly. There are some issues which the machine cannot be expected to calculate, so it has to group train performance into types of performance. The more groups you have, the more important the input data become. You are then relying on a human operator. For instance, if a train has brakes isolated and the human being who enters the data to the system omits to advise of that, then incorrect data fed in will result in incorrect performance out of the system. So we are still down to human factors. Train performance varies widely from one train to another, but train control systems operate on one set of parameters. If they could be made to operate on more train specific data, that would ease the capacity issues.

Cost would seem to be an issue, but that is one for the cost benefit analysis – which I don't believe would say ATP was justifiable. It is therefore a political issue. It is unacceptable to politicians and the public not to have ATP regardless of cost or perceived benefit. If we could do it cheaper, then great, but I don't think that cost is really an issue today. Saying that, if you could make the ATP system integrated with the signalling system, and by so doing make the combined system cheaper to install, the money saved could go to something else. In the long term, that may make a safety benefit.

Qu 8. As an additional point to this, how do you think the trade off between safety and operational flexibility can best be managed?

It is a sliding scale. The fact is that 25 years ago it was quite acceptable to have a couple of fatal accidents a year. People accepted it as a fact of life in the same way that they accepted accidents in coal mines, because at the end of the day life is a risk. Nowadays things have changed, due to public opinion, largely formed by politicians and the press. It is now the public view that we should save lives at all costs in the public transport industry – but not in the road industry where 3000 people a year are killed but everyone just shrugs their shoulders and says" well what can we do?". The fact is that to get the balance right, you do have to first of all take public opinion into account, whether we like it or not, whether it is logical, sensible, fair is irrelevant. ATP is a classic case where public pressure is forcing the industry to install something that is not necessarily the best safety option for the money, and that will have a significant effect on performance.

Ultimately, the only safe railway would be one where everything is at a stand, so it comes down to risk assessment and analysis with a lot of respect for public opinion. That means we have to install ATP and live with the penalties. Whether the balance was really right will be decided by whether it effects the number of passengers that are attracted and the number of trains that can be run by the industry. If performance is reduced by the safety measures we introduce, we will not be able to attract customers. If we are unable to pull in passengers and freight to pay for the running of the industry, then not have the money to pay for further improvements.

Public opinion could be influenced, but the problem is that the politicians and panic merchants, along with those who have a vested interest currently have the whip hand. The public are very critical of the management of the railways as a result of Ladbroke Grove. They are uninformed, and it is the duty of the railway industry to explain better to the public. Unfortunately, the only way that they can do that is through the press, who are more interested in perhaps creating strong feeling against the industry, and creating and manipulating public opinion rather than reporting facts. So at the end of the day, the railway industry somehow has to get to the public with the facts. It is catch 22.

The HSE appears to have taken a defensive stand and gone into self protection mode by getting their retaliation in first. They are therefore publicly very critical of the railway industry, despite being well aware of the pressures that are on it. It should be the HSE playing the leading role in educating the public, being realistic and objective in their assessment of safety issues, and then standing by their decisions. For instance, if a new signalling installation is approved by HSE and an incident then happens, they should say that it is approved and relatively safe, and explain what they mean by that, rather than trying to find some way to blame someone else for somehow contravening requirements. So the HSE should really be campaigning on behalf of the railway industry to get the safety record in perspective. They should be controlling public opinion rather than responding to and inflating hysteria.

Qu 9. What information do you think:

A. Is required by a driver in order to perform the tasks asked of him

In the UK, with route signalling, drivers must have information as to the state of the line ahead – the route set for it, and that the line is clear up to a point ahead to which he has sufficient distance to stop from max. speed. The driver is then expected to know the route speed profile.
For continental driving the requirement is different. He needs speed information as well.
The two concepts require quite different infrastructure and signalling systems. I believe that the UK approach requires the Driver to exercise skill much more than on the continent. We get best performance from our system because drivers, having learned how a train performs over a route, the gradients, speeds permitted etc. can drive the train to its maximum technical performance. The downside is that not all drivers are equally skilled, so there is a bit of variation in performance. I believe that the continental approach tends to bring everything down to its lowest common denominator, because the driver is not much more than a handle pusher. As long as he does what the signal tells him he won't go wrong, but he has little or no scope for making up time. So, in my experience and observation, their schedules seem to be slower, with longer station dwell times to allow for recovery. I do not have a strong view as to which is best overall.

B. Would it be beneficial to a driver to assist him in the performance of his tasks

You could provide something to indicate speed limits to the UK driver, especially to help his route knowledge out where you have enhanced speeds. It becomes far too complex to expect a driver to accurately remember 2 or 3 sets of speed restrictions for different types of trains.

C. Would it be detrimental to a drivers ability to perform the tasks required of him

Anything that might encourage a driver to anticipate what the signalling system will do (like the classic 'approach released signal' scenario) - anything that may invite him to make a miscalculation would be very bad practice.

Qu 10. How do you think that your answers to question 9 would change if the system included ATO?

When it is automatically operating – it is questionable that the driver needs to be there at all. If he is there, I see no need for information other than technical functionality – is the system working properly or not.
If the system fails, them the continental approach probably wins out in this scenario. A driver used to ATO would not have the route knowledge needed to drive to UK practice, but could to the speed targets of continental practice. The only way to use UK practice would be to find a way to keep drivers' route and traction knowledge maintained. Driving manually off-peak may do that. Generally, I think that the speed signalling system would win out in those circumstances though.
I don't think that ATO is advisable unless you have securely fenced-off, vandal proof, out-of-control-lorry-off-a-bridge-proof infrastructure - a tube line or European high speed line that is fenced in and has CCTV, etc.
There is also a problem on main line with the wide variety of different stock types and performances. The ATO therefore has a lot to deal with in coping with the variety of infrastructure, intermixed running, etc. Generally, I think that ATO is not viable on a mixed traffic railway. All trains would need to be fitted, each type operating as close to their optimum performance as the system would allow.
I have never supported the idea that you can't run a railway without a driver at the front of the train, because clearly you don't have to. You can automate it, but you have to make sure that the safeguards match the risks. The advantage of, say, the Victoria Line, is that it is enclosed and contains one type of train, so everything is controlled and uniform. On the main line it is more difficult due to the sheer variety of acceleration and deceleration rates, (and the fact that you can't have a secure railway.) It could be done, but it would need to be incredibly sophisticated equipment to cope with all the variations. If trains cannot stop on sight, no system could operate safely without another way of monitoring for obstructions on the adjacent lines, land slips about to occur, etc. The driver needs to be there to protect other trains as well as his own. You could put CCTV cameras in, but I don't think that would be enough on its own. You also need to keep vandals, livestock and road vehicles out with good fencing, etc. It could be done, but would be much more difficult for a main line than for a suburban line or metro.

Paul Le Vesconte, MIRO (Senior Operations Consultant, Rail Training International).

Questions Raised by D. Woodland on 11/06/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I started on the Railway in 1972 at a mixed traffic depot, called Guide Bridge in Manchester. We had the class 76 1500V DC trains on the Manchester, Sheffield and Wath railway. They were heavy haul freight trains, mostly loose coupled, bringing coal from Yorkshire to marshalling yards at Mottram, Dewsnap and Godley. There they were re-marshalled and taken to Liverpool for export. Later on, we started bringing Australian coal the other way!. We had a lot of other mixed traffic. Oil trains and company block trains (owned by a particular company and operated by BR for them) to Elsmere Port and the oil refineries in the Wirral. We carried a lot of dangerous goods and also had the merry-go-round trains out of Yorkshire down to Fiddlers Ferry power station. We also had a lot of local passenger train work serving the east and west of Manchester, Liverpool and Sheffield. Then there were parcel trains and newspaper trains in the evenings as well.

I started there as 'second man', sat with the driver observing how he drove and having a go from time to time. During that time I was seconded for a while to Longsight depot , which worked all of the intercity passenger trains on the West Coast Main Line at that time. That way I built up my skills base and eventually (after about 8 years) passed my exams to be a driver. I then chose to go to Warrington to take up a driving position. I learnt the route from Warrington to Carlisle North and down to Wilesden South. It was mostly speed freight, running company block trains at high speed (75-80mph.

Then I moved back to GuideBridge, to be closer to home, and remained there for 2 years as a driver. During that time I got involved on the technical side, became a driver instructor and eventually took a promotion as a supervisor into the training school at Crew. I taught drivers how to drive trains, the technical aspects and the theory of how to drive trains, for about 2 years.

I then became a traction inspector (like a drivers policeman) at Crew, where we had a lot of test trains from Derby research at that time. I got quite involved with all of the new build 150 series and push pull on the West Coast Main Line. I then got my first management job as test and projects inspector for the divisional operations manager at Crew. That involved interfacing with Engineers about new build traction units and what they would mean operationally. Taking the engineering problems and producing operations solutions or, taking operations problems and asking the engineers to find a technical solution.

After that, I became Chief Traction Inspector for Central Trains in Birmingham. That role turned into the Chief Standards Manager for the company. I was there for 4 year. Then for 18 months I became the company driver manager, responsible for service provision of drivers, train crew strategy, train crew planning, managing the budget and operational safety. I had 11 depots, 26 supervisors, 5 managers and nearly 700 drivers as my responsibility.

During that time I was seconded to Halcrow-Transmark to go out to Malaysia and train their trainers on a new traction unit that was similar to the Class 323 that we had just got on Central. That was my first taste of looking at a different railway.

When the train operators were bought by the various franchises, I decided to take voluntary severance and went to a company called Rail Training International, who did a lot of work in the far east. I then worked on a couple of projects in Sydney Australia, looking at Driver Only (no Guard) operation of heavy freight haul (10,000 tonne freight trains). I also did some work for KCRC in Hong Kong, for the Athens metro (my first taste of metro operations) and a few other driver training, management and operational safety projects in Australia.

I worked as the expert witness on operational safety and driver issues for Railtrack's solicitors on the Southall and Ladbroke Grove enquiries. I also did a bit on the Watford incident for them.

I then worked for Connex for a while, doing the project work for training 200 drivers for them. Then I came to work on the WCML TCS project.

Qu 2. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

First of all, you need to think about the system that you want to develop or implement. If it is one where you want to change the way that people operate the railway, you need to think about how you are going to manage the culture change. That is the big issue.

I would like to see ATP operating completely in the background and not seen by the driver in any way prior to an intervention. It is important that the driving task is not de-skilled. A system that supervises them in every way would be very oppressive – particularly when they know that everything they do is being recorded as well.

ATP should also give absolute, not partial, protection.

Qu 3. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

Unfortunately, incidents, accidents and public perception. We know where we are with the recommendations made by Cullen and Uff. The public perception and how it impacts on the expediency of ATP implementation has already raised its head.
There is an ERTMS National Implementation Team, being sponsored by the SRA , ATOC and Railtrack. They are looking at ERTMS fo rthe whole network. I think that the main driver behind that was the Cullen and Uff recommendations.

Qu 4. I would welcome any views that you may have as to ways in which Automatic Train Control systems could best be optimised to create the greatest benefit for UK railways in both the short and long term?
There is one thing that strikes me. One of our biggest problems is line capacity and the number of passengers that we can carry at any time. The only reasonable ways to overcome that are to build longer platforms (and run longer trains) or to reduce headways. If ATP could give us shorter headways, that would help a lot.
Also, could ATP offer other facilities, like selective door opening control. That could allow us to run longer trains and prevent doors from opening in carriages that do not fit within some shorter platforms. Improvements like that which increase capacity would be the best form of optimisation.

Qu 5. The terms 'Signalling', 'train control' and 'railway control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system?
Signalling is there to give you a form of train regulation and train protection by controlling the movement of trains, keeping them apart.
Train Control I consider to be managing the operational safety of the railway, manipulating the daily plan to optimise the service, performance and timetable planning. Train reporting, delay attribution, the causes of delays.
Railway Control I would see as a higher level. Regulatory bodies, HMRI, SRA and the rail regulator. Looking at how the railway interacts with other parts of the industry.

Qu 6. Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?
Automatic Train Supervision would be something like an ATP system.
Automatic Train Regulation would be the automatic reporting of trains to signal boxes in real time, about delay, passenger information and assistance to the signaller for regulating trains along their route. Perhaps managing station shunt plans further down the line. It is about managing the railway safely within the constraints of the timetable that you have laid down.

Qu 7. How would you define the capacity of a Railway system (and what would you consider to be appropriate units for your defined measure of capacity)?
It is about headway. How many trains you can get on a given bit of railway at a time, or trains per hour.

You mentioned earlier the potential for increasing the number of people you can get on a train in order to increase capacity.

I do not tend to think of capacity in that way, but now you mention it, yes, it is all about how many people we can carry. Carrying more people in a train could actually reduce the headway between trains, but maybe that is what you want sometimes.
Having dedicated lines for different types of traffic (freight, passenger, inter cities, local all stopper) would be the best way that you could increase capacity on a railway. That is not realistic with the types of railway we currently have. In the future, the use of moving bock may help there.

Qu 8. What information do you think:
D. Is required by a driver in order to perform the tasks asked of him
A drivers prime source of information is currently what he sees through the window. He therefore needs to understand the meaning of the signals that he sees. The signals must be unambiguous to him.
Because we are in the UK, operating a route signalling rather than speed signalling system, he also needs an intimate knowledge of the route and all its speed restrictions. The information on speed restrictions is repeated by lineside signs in order to remind him, but he doesn't need that.
The art in driving a train is stopping safely where you want to stop and keeping to timetable. So, the speed the train is running at and a measure of the brake force available is necessary. That doesn't mean 250 tonnes of brake force. That would not mean much to the driver. You need to tell him how many vehicles the train has. The driver can relate to that mentally.
When the driver is braking the train, he starts at YY say, he has a course idea of where he wants to be and will brake to be around 60mph by the Y aspect. It does not matter if he is 5mph above or below that. At that stage, it is a rough calculation. As he then approaches the signal where he has to stop, his calculation

becomes more refined. He then starts comparing the brake cylinder pressure, distance to go and speed, refining his brake application as he gets nearer the signal. So, it is brake, speed and distance to go that he needs.
I think that is all of the basic information he requires.

E. Would be beneficial to a driver to assist him in the performance of his tasks
Gradient information, badly sighted signals, tunnels, etc.

C. Would be detrimental to a drivers ability to perform the tasks required of him
A proliferation of speed boards and other lineside information can become a problem

Do you think that the information required by a driver changes at all when you have ATP?

The way I see it, no. Whilst we have signals acting as the prime source of information, we should promote head up driving at all times – even at speeds above 125mph. The information on the DMI should therefore be kept to a minimum in order not to distract him.
When we get to moving block with complete in cab signalling, there needs to be a completely different philosophy. The DMI becomes his primary source of information.

Do you think that changing from lineside signals to in-cab and back would be difficult for drivers?

I drove from Paris to Lyon on TGV. About 15km out of Paris you leave the conventional signalling and enter a cab area. I found it uncomfortable at first, I still wanted to take train running information (signals) through the window. The French guys did not seem to have any trouble at all. After about an hour, I had settled into it. On the way back, I had no trouble at all with the transition back to lineside signalling. So, I don't think that it would be a problem.

I think it is the TBL system that provides in-cab information on temporary speed restrictions and stopping targets, but deliberately does not provide in cab information on permanent speed restrictions. The idea is to encourage the driver to keep looking out of the cab window in order to be aware of where he is. Would you support that sort of approach to 'minimum' information?

On a route signalling system, yes. On a speed signalling system, I am not so sure, since you tend to have set speeds for crossings and things anyway.

There are a lot of people who think that speed signalling is the way for the future, particularly with ATP systems (which effectively work on the basis of speed supervision). Do you have an opinion as to whether one is better than the other?

I have operated all of my life on a route signalling system, which makes me a bit biased. Obviously, there are advantages to speed signalling. Driver route knowledge / training costs and infrastructure costs can be reduced (by standardising crossing speeds and associated interlocking). From a cost / benefit point of view, it is probably the way to go. Considering that the rest of Europe uses speed signalling, I think that it is inevitable that we will change over eventually. The operators in the UK will be dragged into it kicking and screaming! It will probably come on dedicated lines at first.

Qu 9. How do you think that your answers to question 9 would change if the system included ATO?
There would be a fundamental change. What is the driver then doing there? It is really to detect trespass and anything foul of the track – but at high speeds, he is unlikely to be able to do anything about that in time to avoid a collision. When the driver is left in the cab, like on the Victoria line, he is only really there as a comfort factor.

If a driver is required only in case of failure of the ATO, what information would he need when he is driving?

He would need the same information. Under that situation, the problem is competence decay. You therefore need to maintain the competence by regular training.

Do you think that could be managed by manual operation off peak and automatic operation in times of peak capacity demand?

Yes, that is a fairly good compromise solution to maintaining competence. I would have to think about how the driver would respond to effectively being redundant and only allowed to drive in order to 'keep their hand in'. At the moment, there is a lot of job satisfaction in driving a train and being in control of getting people from A to B. With ATO, you may loose that feeling of reward, which might de-motivate the workforce.

Qu 10. The introduction of higher speed operation, higher capacities or changes to track layouts may all offer operational benefits, will influence the commercial viability of projects and will also affect safety. How do you think the trade off between safety and operational flexibility can best be managed?

By having people around who understand the limits and parameters around the system you are working in. The kind of sense check from experienced operators and engineers can be invaluable.
Running risk models against what you want to develop is also an important tool.
You need to consider wider issues. How is it going to effect the operator, the cleaner, and the passengers.

Supplementary questions raised on 24/6/2002:

Qu 11. I am interested by the use of warnings prior to brake intervention on BR-ATP and the proposals for TASS and TCS. What do you see as the benefits of such warnings during ceiling speed supervision?
Warnings are nice to have, but not essential. As a driver, it is nice to understand where your thresholds are. This makes warnings particularly useful for braking curve supervision – especially under abnormal conditions such as reduced adhesion assumptions or isolated brakes. Then the driver needs some advice on where to start braking to avoid an intervention. For ceiling speed supervision, I do not see such a need. You may want to talk to a human factors expert about that.
Modern standards require speedometer accuracy of ±2mph at 100mph. The driver may need another ±2mph on top of that to effectively control the train's speed (the acceleration and brake control may not be fine enough to do better than that). On that basis, intervention without a warning would seem fair as long as it did not occur below 4mph above the permitted speed.

Qu 12. If a warning is going to be given, it seems sensible to me that the driver should be allowed sufficient margin between warning and brake intervention speeds to allow him/her time to react and bring the train's speed under control under all reasonable situations (accounting for likely acceleration due to both traction and gravity during the reaction time and any traction cut off / brake build up delays). What are your views on this?
There would be no point in the warning if that were not the case.

Would it be reasonable to assume that a driver making an error that causes the permitted speed to be exceeded could realistically still be accelerating at full rate?

Driver training teaches that you should accelerate the train at its maximum rate up to the permitted speed. This is necessary if the timetable is to work. In practice, drivers will accelerate at maximum until they are about 2 to 3mph below the target speed. They will then ease off gradually to bring the train to the desired speed. As they only ease off relatively close to the target speed, I would think that it is reasonable to assume that the train could be accelerating at full speed when a warning is given.
It is interesting to note that diesel trains cause more problems than electric ones when it comes to cutting traction off. The driver has to close the power handle fairly slowly in order to avoid problems of the engine racing. The throttle is controlled by an engine governor that receives power demands from a load regulator. If the load is removed too quickly, the engine races because the throttle can not be cut shut off quickly enough. That can lead to engine damage. In order to overcome the problem, trains are fitted with a trip device that will bypass the throttle if the does have to reduce the load quickly in an emergency. The trip device requires the train to be brought to a stand and the driver to go back into the engine compartment to reset it. Hence it is not suitable for use in service braking.

Trains accelerate much better at low than high speeds. What would you think was the minimum permitted speed that a train may encounter?

On approach to the buffer stops in a terminal station the permitted speed would be only 5mph.

What gradients would you think it was reasonable to assume may exist during a brake application following receipt of an overspeed warning?

On the West Coast Main line, the worst gradients are around Euston. At Camden bank it is 1 in 77. Between Euston and Camden the gradient reaches 1 in 70. Also at Shap it reaches 1 in 75. You would need to consider gradients that steep at least. There are higher gradients elsewhere on the network.

Supplementary question raised on 11/7/2002:

Qu 13. Would normal driving practice ever permit operation at the railway's theoretical headway?
A driver would normally start to brake at a YY aspect.
In practice the exact response point will vary according to the type of train and the drivers feel for its response. For example, with a tap changer loco, the driver would probably start reducing traction at the green signal before the YY if he can see that there is clearly a train in front. This is because it takes some time to reduce power before the rheostatic brake can be demanded. The driver would then apply the brake if the signal is still YY when reached. If sighting is not good enough to allow this, the driver will cut traction and apply the brake at the same time – but that will mean that the pneumatic brake has to be used. With modern solid state traction control, the driver does not need to cut the traction so early. If the driver

then sees that the next signal is still YY, he will ease back further, trying to match his own speed to that of the train in front. He will keep doing this until he finds that he is running on green aspects. After that, he will regulate his speed to try to keep on greens. He will probably see when the train in front is getting further away by being able to read through a few signal sections, and will then accelerate up again.
The driver does this to save on brake / traction rate changes, making the journey smoother.
On top of that, the operating companies all now advocate 'professional driving techniques', which offer guidance on how to approach a stop signal. Each TOC has their own version. You would benefit from looking at a few of them.
Basically, the drivers will never get to the theoretical headway in normal operation. They will always try to drive with a greater separation.

When you consider the headway effects of an ATP system, it is normal to work out all of the system and driver response delays, and to compare the theoretical headway permitted by the ATP with that of the signalling systems theoretical headway. On that basis, the headway permitted by an overlay system always comes out worse than that of the conventional signalling it overlays –sometimes a lot worse. However, I suspect that this is not a fair comparison, since under conventional operation (particularly with 'professional driving') the driver does not achieve the theoretical headway performance under conventional operation anyway. The ATP only has an impact in reducing capacity if it warns or intervenes at the speeds the driver would actually be doing.

That is right. When you consider the headway effects of an ATP system, you only get a meaningful comparison if you compare the warning and intervention speed / distance curves with the drivers performance speed / distance curves, rather than the theoretical limits of the signalling system. The ATP only has an effect if it causes the driver to slow down more in order to avoid interventions. I think that the effects of ATP when compared against actual driving practice will be far less than is currently being predicted.

Jerry Lewis (Principal Systems Engineer, Alstom Signalling Ltd).

Questions Raised by D. Woodland on 15/02/2001.

Qu 1. I understand that you have been involved in the development of a number of ATP systems for use both in the UK and abroad. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of what your involvement was in these projects?

I was involved in the Chiltern ATP scheme and the KCRC scheme. The Chiltern Scheme involved application of ATP on top of existing signalling , whereas KCRC was associated with the complete re-signalling of a railway. On the Chiltern Scheme I developed a holistic understanding of the system and specified the manner in which it should work. Then at the end I specified a number of design rules for applications. On KCRC, I initially led the scheme plan design, and then moved into a similar sort of a role to the one that I had on the Chiltern scheme. This involved identifying problem areas, proposing modifications and assisting people to develop ways in which they could be put into practice.

Qu.2 The terms 'Signalling' and 'train control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control system'?

I think that people use these terms in lots of different ways. I would say signalling was a subset of a train control system, in that it is a means by which information is conveyed to the driver. It includes, of course, interlocking and such like. I would also include the higher-level functions like supervisory functions, train supervision functions (IECC for instance) within the definition of "signalling".
"Train control" includes things like automated train operation, which I would not consider to be part of "signalling".
I would include everything that is used to control trains in "Railway Control Systems". On a Metro system I would want to include things like fire protection, crowd control, lift control and escalator control, because there would be a need for certain links between the control of those things and the control of the trains. On a less intensively trafficked railway, these links wouldn't be anything like as important. Therefore I might not consider them to part of the Railway Control System, but as separate entities.
There probably is a rationale in using slightly different definitions in different contexts, where the importance of what you include within the core system and what you include in separate sub-systems could well be different. I think somebody with a metro background would think differently to somebody with a main line background.

Qu.3 Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

Train regulation specifically refers to controlling of the trains in order to optimise the service, at a higher level than the signalling. Maybe to hold a train back in order to even out the intervals on a metro, or to hold a freight train in order to let a passenger train through first on a main line. Coasting to optimise energy consumption would also come under ATR. Basically, anything that might be expected to happen on a regular basis in normal traffic in order to optimise the service. Regulation occurs generally in line with a reasonably established set of rules.

To me, supervision refers to the entire process of the high level control of railway operation. ATS is a superset of ATR. To me it includes ATR, but would also include recording of information for future analysis and watching for things that arise that might need to be done. Turning trains back to fill in the service, cancelling trains and other higher level decisions would be a part of ATS. Basically, anything that happens on a less regular basis, when the intended service can not be maintained and high level decisions on how to change that service need to be made.

So how do regulation and supervision relate to train control?

I would include supervision within train control. The whole business of controlling trains, right from the point at which the train is available to leave the depot, making all decisions of what to do with that train and implementing them right up till the time that it re-enters the depot are part of train control. ATR is a part of ATS which is a part of ATC.

ATR and ATS, are functions which are not safety critical, although they might be safety related in a tenuous way (due to platform crowding and delay implications). Other systems are supposed to deal with the safety factors, and the risk of them failing is small. All you are doing with supervision functions is stretching those systems a little bit more than we would have done if the service had been working properly.

{Clarification given on 16/5/01}

Historically (10-15 years ago), we would have considered ATC = ATO + ATP, based largely on the products then available. We then had a supervisory system (or ATS) with the interlocking below that and ATC (including ATO and ATP) below that. The Central Line, Victoria Line and RECEFE fit this pattern. Now we have blurred the distinction in order to get improved performance. I think SACEM is a good example, as is TCS level 2. With these new systems, we have to redefine the way we view the system, and we come up with something like this:

Qu.4 How would you define the capacity of a Railway system? Please state what you would consider to be appropriate units for your defined measure.

I think that depends on your railway system. On a metro, the important thing is that capacity needs to be there when you want it. You are looking to carry people, and you need to schedule maintenance around that. Therefore, I think a suitable measure might well be passengers per hour.

If you are thinking about heavy-haul railway, then whether you have capacity exactly when people want it or not is less important. If the railway ran six days a week and you did all of your maintenance on the seventh day it wouldn't matter. Or you might, for example, run trains four hours at a time and do maintenance in the intervening hour. What you want is the aggregate haul over a period of maybe a day, a week or maybe even a month to be what you need. Therefore perhaps tons per day or tons per week may be more appropriate.

Those are the two extremes perhaps, but they show that you need different measures of capacity and different time periods in which you measure that capacity to make it meaningful. I suppose for the signalling or railway control engineer this distils down on both types of railway to what we do now, in terms of talking trains per hour. However other constraints also operate to determine the size of the train, which overall capacity also depends on.

On a mixed traffic railway, you again have a different problem because you've got different kinds of traffic. I would probably be happy with the trains per hour type of measure on a mixed traffic railway providing you then define things like maintenance periods. What you've done to get to that though is break down the high level requirement - those being tons per week or passengers per hour... or whatever. You've broken them down by defining the size of the train, the number of tracks and all those other things to end up with trains per hour capacity. So yes, for the signalling engineer, I think that trains (of a specified type) per hour is probably a good way of defining capacity , but at a higher level it's probably not.

The problem with the trains per hour approach is that it is point specific, so on most railways the actual capacity as defined by trains per hour varies up and down the line.

They would yes - and the need would vary up and down the line as well. The capacity of a simple line between two points is the capacity of the worst case point on the line. I think that, probably, the more difficult thing would be defining junction capacity, because that requires you to be clear about what you need (which way you go through the junction and how many paths you want through that junction between each of the possible entrances and exits). I find it difficult to see how you could make it meaningful in anything other than trains per hour with a specification of the traffic pattern. Perhaps junction capacity could be specified by first defining a cycle of trains required to pass through the junction and then specifying cycles per hour. Something of this nature could also be used on a mixed traffic railway.
On a more complex railway, capacity of the entire line could be affected by interactions between junctions, perhaps some considerable distance apart particularly if a journey time between the junctions is specified.

Qu 5. Given the low costs of TPWS as compared with a comprehensive ATP system, do you think that TPWS can be seen as achieving the optimal ATP solution in terms of cost / benefit, or are there other possibilities? Please explain why you hold the views that you do.

An optimal solution is very much a matter of judgement and I suppose, to a large extent, it depends on what you can afford. Clearly a TPWS system fitted to 99% of the railway is much more effective than an ATP fitted over about 10% of the railway. I do see TPWS as having problems/shortcomings. I also feel that, if TPWS became common, when some incident occurs that could have been stopped if TPWS were modified in some way, TPWS would get modified. In this way the TPWS system would then grow, possibly becoming something of a mess. I think ATP could give a better long-term benefit - it would be more amenable to the developments that we are likely to want. For a start, TPWS is not effective above seventy miles per hour, and we already have the idea of TPWS+ to extend that speed range. So if you want to increase the speed of a stretch of the railway, are you now faced with turning TPWS into TPWS+ or even TPWS++?

As things stand on the UK railway at the moment, TPWS is being fitted on critical places at relatively low cost (but only at critical places) *and predictions have been made that it will prevent 70% of all ATP preventable accidents. On that basis, do you think that it would be financially justifiable to fit the more expensive ATP system in order to stop the remaining 30% of ATP preventable accidents from occurring?*

Well that depends on whether you're simply thinking of now, or whether you're thinking about the long-term. Let's assume that the TPWS figures are right – will they remain right in the context of what's likely to happen on the railway over the next ten or twenty years? I think probably that they will become less effective. You could also argue that ATP could be installed only at critical points - it would be more difficult to make that sort of argument because whereas with TPWS you only need to fit one or two signals to protect a critical point, with ATP you would end up fitting a number of signals to protect the same critical points. But the cost for trackside equipment, if we're talking ERTMS level one, would clearly be lower than trying to fit the entire railway (which is the way we are going). The other thing is, of course, the high costs of train fitment with ATP – it would be interesting to see whether that could be reduced in some way, without heavily affecting the functionality of the system. Cost is obviously a big factor in whether it's economic to fit ATP but, in any case, even if you decide it was economic at the current cost, it would still be sensible to reduce it if you could wouldn't it?

Whenever you reduce costs you tend to incur some dis-benefit in proportion to the amount you have saved. If you follow that process downstream do you get to TWPS as an ATP system with minimal costs for maximum benefit?

To say TWPS as minimal cost is a bit of an inaccurate statement – cheaper, but not minimal by any stretch of the imagination! But yes I take your point; you could end up reducing the functionality. I don't think that's necessarily the case though (this is purely a personal opinion of course). As our technology changes then we would have to see what was available wouldn't we?

Qu 6. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

If you mean installation, testing and commissioning, Generally speaking, it doesn't appear too difficult to fit ATP onto the railway, in terms of actually getting the system in and working. You can install equipment bit by bit and there are fairly clear-cut interfaces between conventional signalling and ATP in terms of current designs. I therefore think that is not too serious an issue - or at least it's one that has been well thought about and perhaps not one that needs to be pursued. Trains tend to be more difficult to fit.

Previous experience indicates that it is not something that can be accomplished while the train is in the depot overnight, which results in the train having to be removed from service while it is done.
If, on the other hand you are referring to the more fundamental issues of making the ATP work safely, this is going to depend on the system but problems are likely to arise in making sure that it treats all of the situations that are likely to arise in a manner which is consistent with the underlying signalling system
I think probably the biggest problem is the impact on capacity which is something that needs to be considered in depth if the line is busy. That is the possible decrease / increase in capacity. A second thing is it's affect on the drivers (whether it impedes them or not). There is, I understand, evidence that drivers who are driving with ATP have more SPADs than those who are driving without. I'm not clear on whether they have these SPADs when they go off equipped lines or they are ATP protected 'little SPADs' that stay within the. So perhaps that's an issue that needs to be investigated

Any idea where I might be able to find data in support of that theory?

Well, a Railtrack representative has said it to me. He's also said that Railtrack are going to do an investigation into the performance of the Chiltern and Great Western mainline schemes starting sometime round about now. I'm not quite sure who it is. Talk to Steve Roberts or somebody who could put you in contact.

Qu 7. I would welcome any views that you may have as to ways in which ATP could best be optimised to create the greatest benefit for UK railways in both the short and long term?
I think we have the capacity problem. As far as possible, we want the trains to as they would without ATP, or preferably better. Now to talk in terms of optimising ATP is perhaps not right It may be better to talk more in terms of optimising the railway. For instance if you have an ATP fitted railway, could you reduce the margins between authorised speeds and dangerous speeds? So you might now be able to say 'I leave a margin between the way I maintain this curve and the running speed, but I have an ATP fitted railway where I can be confident that trains won't exceed the authorised speed. Maybe I can now increase the allowed running speed, or reduce the maintenance'. That's the sort of process one might be looking at. Ultimately your use of the ATP system rather than the ATP system itself – yes you've got the whole railway, not to simply look at ATP

That sounds like a longer term thing, because would have to rely on all trains being fitted and working with ATP?

Well, you say that – you could have separate speed limits for fitted trains and non-fitted trains. Maybe on the fitted trains you could then put the information in the cab to avoid additional lineside signs.
I don't think you could use ATP to shorten braking distances or anything like that because they are what they are, but if you group ATP with a cab signalling system then there are potential benefits from doing away with the line side signals and all that comes with that. I'm sure the potential savings there are quite large. If you do away with the line side signals, that would require all trains to be fitted, but you could have a reduced signalling system which would deal with the odd non-fitted train – which would just handle a low capacity service.

As was planned for the JLE.

Yes, and in fact we proposed it for the Northern Line as well. I feel in the longer term that a level 3 system (doing away with track side equipment), really is the way to go if you think in terms of optimising existing ATP technology. Clearly optimising the technology itself is another thing.
In previous systems the braking characteristic used by the ATP system has been extremely conservative. It would be useful to see what benefits could be achieved without significantly increasing the risks by making it less conservative.

Qu 8. As an additional point to this, how do you think the trade off between safety and operational flexibility can best be managed?
Not an easy one is it?! Railtrack has its' safety management system at the moment. It relies on risk assessment and assessment of benefits – the way in which this sort of thing is approached nowadays is, if you can relate the risks and the benefits 1 to Pounds, then that gives you one way of doing it, albeit very crudely.

What quite interests me is there is a lot of things you could potentially do which would come out great in economic terms: maybe things like higher line speeds or operating within relative braking distance, where you introduce new finite risks (although they may be small) of an incident happening that would result in damage, casualties and fatalities, on a potentially large scale. How do you think things like that can be handled?

This is normally very difficult because it's not something you can do within the industry. It's something that the wider society has to buy into and probably doesn't understand (or doesn't want to understand). But clearly if there are benefits then it would be useful to find an objective way of measuring those benefits against the risk. What I would want to consider are things like other risks which are diminished by doing these things. For instance, if, by operating a relative braking distance, you avoid overcrowding on the platform, then yes you have increased the risk of a collision (maybe), but you've decreased the risk of somebody falling off the platform and being run over! You might have even decreased the risk of having to shut the station or people spilling over onto the road and getting run over by road traffic! It's not an easy thing to do is it? I don't think it would be accepted by the general public, even if you could show the economic benefit, unless you could show that there was a decrease in risk in other areas. But having said that the increase in throughput has got to be attractive and perhaps the difficulty there is do the public see it as being an increase in profits for somebody? Or a big decrease in their fare?

Qu 9. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

I think the thing is controlled very much by government legislation. I don't really think that Railtrack or the industry as a whole would want to fit it if it didn't have to (unless the cost came down considerably). To encourage wider, quicker application, I suppose the first thing is HMRI pressure or legislation. The second thing is to reduce the cost, either of the ATP system or some of the other costs. For instance as I suggested earlier, maintaining the line for a lower speed because we're now sure that we can stay within the envelope (which we weren't sure about before), or increasing benefits by maybe running faster round the curve than we could before).

I suspect that the financial case is much more difficult to make now that the industry has been split up. Railtrack has to bear most of the costs and problems associated with ATP, but I imagine that the TOC is likely to bear the costs of an accident caused by driver error.

It would be interesting if you took out the costs in terms of death and injuries, to see how things stack up in terms of cost, lost revenue, damaged vehicles, damaged track.

Yes, I have done an estimate of that. For each life lost in the accident, the accident's likely to cost you seven million pounds in terms of lost revenue and repair costs, which still doesn't stack up to the cost of installing ATP, but It is significantly higher than the criteria that tends to be used.

So that, therefore, means what you need to do is predict where your accidents are likely to be. Which brings you back to TPWS solutions or something similar.

Which is why I was wondering if it was a cost optional solution.

Well what you must recognise is that even at the locations where it's fitted, TPWS doesn't guarantee safety, (neither does ATP necessarily, although it does better). But you won't guarantee to prevent every accident at the location in question so that's not the whole story. What that does suggest perhaps is better ways of understanding risk. Maybe there are other factors involved as well which are less obvious.

Qu 10. How do you think ATP could best be integrated with the rest of the control system?

Well of course at the moment we're not, nor did we really on KCRC – well it depends on how you mean integrated. In equipment terms in the case of KCRC the ATP equipment just sat on top of the interlocking. However, in thinking terms what the ATP did and what the interlocking did was integrated. For instance shunt routes were treated very much like shunt routes on TCS – you just supervise it to a maximum speed and if you run past the balise you trip the train. But we then did consider how long the overlaps were beyond those balises to make sure that they were consistent with the speeds we were able to run in shunt mode. We were only able to do that, of course, because we were replacing the interlocking at the same time as putting in the ATP. So we integrated it in that way, even though physically, it just sat on top of a separate load of boxes. Clearly there is scope to integrated it thinking wise – I hope that's what we'll be able to do on TCS within the constraints that we've. Unfortunately, of course, we are impeded in that by the fact that the TCS-Ready specification never really had enough information. In the longer term, of course, if you're thinking of level 2 (or even level 1) you could physically integrate the equipment. With KCRC the boxes were designed to have serial links between them (between the interlocking modules, TFM's and the encoders), so there was a definite intention that they work together. What would be interesting would be level 2 or level 3 to provide an interlocking function and an ATP function in the one part of the same software wouldn't it? – on one computer.

What do you see the advantage of that being?

I think it would reduce costs. At the moment we are transmitting everything in the form that you want for line-side signalling and then turning it into the form that you need for ATP. It would seem to me that logic says why bother with the intermediate step, it's not necessary. If you're not constrained in the way you transmit by conventional signalling, there must be an advantage because you could reduce costs. How big an advantage I don't know – if you think about the level 1 solution we have so far– think of the costs of equipment, installation to make those bits of equipment work together. Installation which is all hard wiring, so it all needs to be wired and tested. If you think the West Coast Mainline problem it's all going to have to be wired out in the field, tested out in the field – there must be big savings if you can integrate those functions. You can probably integrate even more, if we use our imagination, we could significantly increase the amount of integration.

Qu 11. What information do you think:

A. Is required by a driver in order to perform the tasks asked of him

Well essentially you need to know the limit of movement authority and the maximum speed at which he can travel at his current location and up to some distance ahead, depending on what the changes in line speed are. If you're thinking of requirements, that's about it.

B. Would be beneficial to a driver to assist him in the performance of his tasks

That's more interesting because clearly the driver can do other things in terms of optimising the way the railway runs by making judgements on the way signals clear in front of them and, knowing bits about what the timetable is, make an assessment of what's going on. He can therefore try to optimise the way he drives the train in line with these judgements. Now clearly there is a possibility of giving him some more information which could be in the form of what the preceding train is, how far ahead it is, how fast it's going, where it's going to diverge, but all that seems rather complicated. Although that is the sort of thing the driver is using today, but again it's not things he's told, it's knowledge that he's got and I suspect that the information he's got only allows him to perform that optimisation fairly poorly. So it would be nice to think that we could give the driver something more which could be done by deriving/predicting speeds of the previous train, predicting when it will clear from in front of him and giving him some information based on that sort of thing. There is, I suggest, a significant potential to benefit from that, both in terms of energy consumption and in terms of optimising the capacity of the line.
You could give the driver an indication of the speed that he ought to be driving at a particular time, not purely on the basis of how fast it is safe to run. You have got a speed that it is safe to run at, but you can then have another speed which it is desirable to run at to optimise the performance of the railway – it clearly cannot be any higher than the safe speed, but could be lower .What that speed was would depend on a lot of factors – whether you're looking to optimise capacity, or whether you're looking to optimise energy consumption, what else is going on at any time and it would depend on the types of train. It certainly seems to me that if you've got heavy trains which accelerate slowly, there is likely to be a very real benefit in getting the train to run through a junction without slowing down, rather than pulling it up and having it re-start. So you tell the driver to slow down well back.

An interesting thought. If we had a display to the driver telling him how fast he should be going at any point in time, that's likely to be something that fluctuates by the optimal conditions quite a lot. Therefore, he would have to pay close attention to it. Do you think it would have human factor implications in terms of watching his target speed to make sure he's meeting it rather than looking out through the window?

Well we're already providing the driver with a speed that he should be travelling at for safety purposes, if you look at the ERTMS Spec. – it doesn't necessarily mean that you apply it that way but that's the sort of thing we're coming to. I would have thought that when you're in a braking curve, that is going to change very much more frequently than a speed given to you to optimise the performance of the line. I think also we should aim to be in a position where we give the driver varying information and tell him when it changes so that he can then look at the display. That's obviously a bit more difficult because you've got to consider something which is continuously varying. In fact you could make it a speed – this is the optimum speed you're driving at now, which would take away the drivers judgement, but probably give a better result because most of the time the system would have a lot more information than the driver could use so it could calculate more accurately what needs to happen. The driver is, most of the time, using inspired guesswork and using his skill to decide what to do with that inspired guesswork. Using the accurate information then you'd probably end up with a better job.

C. Would be detrimental to a drivers ability to perform the tasks required of him

Well you just mentioned the possibility that continuous information might distract the driver from looking out the window. It might not be a bad thing for them not to look out the window, of course, if you were on a railway that you had designed on that basis. The point is why would we want him looking out the window? – because he can look for things that are not right about the infrastructure, people beside the line, people working on the line, ... if you were to supervise by CCTV, for instance, the entire line perhaps we wouldn't need the driver to look out of the window any more. Perhaps we don't even need the driver to be on the train any more. So it may not be such a bad thing. What I think would be detrimental to a driver to do his task is to do things that are distracting, the worst possible thing has got to be information which is not accurate. I therefore think it is important that any ATP system is designed with information that is accurate and if at any time, we can't be sure, then this must be very clear. I'm not sure that conflicts between information in the cab and what's on the line side are detrimental in the way that a lot of people seem to think, as long as both sets of information are safe and there is a proper strategy in place for dealing with the conflict. I don't feel as worried about it as many people do.

I assume you mean by that if the signal is displayed red and the cab display saying it is alright to proceed, or vice versa, then there needs to be some kind of strategy that tells the driver which to follow.

Right, obviously it's a reasonably simple strategy and in TCS level one there are likely to be places like that – well I say that, but it depends on what information you give him doesn't it? One thing I would certainly like would be for the driver to have information which says that the movement authority exists, even if we don't call it a movement authority. However, information in the cab that says that there is a movement authority when there no longer is (which is a distinct possibility)would be detrimental. The other thing that might be detrimental is if we have too many things happening at the same time. I'm a bit concerned about the idea that you've got a lot of systems in the cab all doing things simultaneously and the driver's expected to manage them all – it's a workload issue. To give the driver too much information, his workload becomes too high.

Following on from that you pointed out that you might be able to do various things which mean you don't really need the driver looking out of the cab, or even need a driver at all.

Qu 12. How do you think that your answers to question 11 would change if the driver were an ATO system?

I don't think it makes a lot of difference really. Clearly you need to know what the movement authority is and what the speeds are. It's possible you might be able to cope with more information and the issues about how you present it wouldn't be different. With the driver there, then you obviously have to worry about presenting it in a manner which the driver can interpret easily. If it's ATO we design an interface to suit both sides of the coin. We could probably give the train very specific information about what it was required to do because one of the big things that we could do (if we did the things I suggested in question 11b is make the driver a bit of an automaton. If it's an ATO system it isn't going to get bored so you could give very detailed information, very specific information without having to worry about issues like that. But of course you lose the driver's ability to make judgements on situations that you hadn't foreseen.

So, is there any other information that might be needed for ATO operation which you wouldn't need if you got the human driver there?

You mean on the control system? Obviously there are things which you need to deal with. You need to understand how you're going to deal with people on the track, if you have 'red zone' working you have to deal with how you warn people (the driver does play a part in that today). You have to deal with the issue of how you're going to monitor the track, issues associated with starting from stations (to whom you give the right away has changed -maybe you won't need to bother about giving the right away in the same way as you do today because you have platform screen doors, you treat it almost like a lift maybe. What other information might be needed? – there are endless possibilities, there are possibilities such as warning of rail adhesion, and the trains all passing information up and down the line about the adhesion conditions in various locations , thus enabling you to take advantage of the conditions of any adhesions which is good, without having trouble when adhesion is bad. You could also pass information about other things that are going on between trains.
Interview with:

Interview by e-mail with:

Sam Macano (Principal Engineer, Alstom Signalling Inc., USA) and Jim Hoelscher (Staff Engineer, Alstom Signalling Inc., USA).

Questions Raised by D. Woodland on 29/10/201 to Mr Naor Wallach (Director CBTC Systems, Alstom Signalling Inc., USA). Responses were given by S Macano and J Hoelscher on Mr Wallach's behalf on 08/02/2002.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

Sam Macano -I have worked for ALSTOM Signaling (previous General Railway Signal Co. and Sasib Railway) for over 35 years. I am currently a systems engineer working on the Canarsie Phase 3 Project; a communications based train control system for rapid transit. Most of my experience has been with carborne related systems (Cab Signal, ATO) or with non-safety related wayside railroad systems (yard automation systems or non-vital train to wayside data communication systems)

Jim Hoelscher – I have been with GRS/ALSTOM for 31 years. I a currently the systems engineer on the Advanced Civil Speed Enforcement Project (KVB based system) for Amtrak. Most of my career has been in product design engineering and most of that has been in the area of AF track circuits, cab signal, and VPI (interlocking control).

Qu 2. During a review of literature about railway signalling and control, I came across a quote by an English author (Mr P Connor) stating that *'Time interval working is not common now, but is still in use in the USA, where following trains in the same direction are authorised through a train order system to follow each other into an un-signalled single line based upon a time interval separation (usually 5 minutes). When this system of working is used, any train failing to keep to the normal speed of the section is required to drop a flare that has a five minute burn time. A following train is not then allowed to pass the flare whilst it is still alight. Similarly, if a train stops, the conductor is required to walk back a safe braking distance, place detonators and show a red flag or light to following trains'.* I would appreciate confirmation that this is the case (or correction if it is not).

I am not familiar with such systems in use today. If they are used I suspect they are only used in short haul local railroads. If a time-based authority is used, I think it is on the order of hours rather than minutes. Most non-signaled "dark territories" used radio based train orders with data radio and on board printers for confirmation.

Qu 3. I would be interested to know how you define vehicle-based control?

In the context of a system where the location of other trains and the state of wayside appliances (point machines) is not directly controlled or determined by the onboard equipment (that is, this information is conveyed to the vehicle via data communication means), a vehicle based system is one in which the vehicle itself must determine if it is operating within safe parameters (train speed as a function of location and its distance from known limits of authority – such as a train ahead or the entrance to an interlocking without an assigned and locked route). A supporting subsystem must be used to allow the vehicle to determine its location with respect to the guideway (track) infrastructure. This may be a beacon type subsystem or a radiolocation type subsystem such as GPS.

Qu 4. Most transportation modes (road transport, aircraft shipping) rely in varying degrees on vehicle-based systems to ensure safe control and avoid collisions. In contrast, railways rely on the operation of infrastructure-based equipment to set up the correct path at junctions between origin and destination. There have recently been a number of papers in the UK noting that technical factors of infrastructure based systems (including system delays and common mode failures) are now the main limiting factors to increasing rail system capacity. I would therefore welcome your opinions as to the benefits to be gained from the use of vehicle-based control to overcome these limitations?

I strongly disagree with the above statement. Ground based air traffic control is the <u>only</u> method used. Collision avoidance systems on board an aircraft are usually short range systems for emergency situations only. I believe that traffic signals are the only method of avoiding roadway collisions between intersecting routes. Shipping is the only transportation mode where this may be true – and this only because the speeds are relatively slow

- *what do you believe the potential to be for reducing system delays?*
  I believe the potential is minimal for railroad systems. A well designed Centralized Traffic Control System does not induce significant system delays. I believe that most delays are due to train breakdowns or infrastructure problems (broken tracks, etc.). There may be some advantages in rapid transit systems where closely spaced trains are commonplace. In my experience wayside based route selection is usually not a limiting factor in train spacing. The limiting factor is the granularity of train location (position and train length). The finer the granularity, the closer the train spacing.
- *what other benefits may arise from vehicle based control?*
  There may be some benefit to a vehicle based route request in transit systems where diverging and merging routes are commonplace and train spacing is very close. A vehicle based train control also has the capability of obtaining a very good position resolution and is more cost effective than wayside based systems. Also this type of control system is readily adaptable to existing systems and is very flexible in the control functions that can be implemented.
- *What functionality do you think could realistically be located on trains and what do you see to be the potential mechanisms for achieving this?*
  Train based control has already been achieved on several systems. On board train localization, on board databases with track infrastructure characteristics, on board dynamic speed profile generation and speed enforcement, and continuous train to wayside data communications have already been achieved. The functions that would be difficult to achieve safely on board the train are; 1.)On train interlocking of conflicting routes, 2.)location of other trains that are in the

forward path and/or on conflicting routes. The ability exists to control the position of wayside points and obtain their status, but problems occur when this control must be safely co-ordinated with the activities of other trains.

Qu 5. I would be very interested in details of any vehicle based control schemes currently in implementation or development and the equipment that these schemes are based upon.

Our current product development is a Communications Based Train Control System (CBTC). We consider this to be a vehicle based control scheme. The system is a moving block system. The following is a description of the characteristics of this system.

Vehicle Functions

- Localization capability. The system uses trackside mounted encoded beacons (balise) that contain an identity code. When a train passes over the beacon it reads the encoded information thus allowing it to determine (in conjunction with the onboard database) where it is. Beacons are spaced at random intervals on the tracks. The train uses tachogenerator information to determine its position when it is between beacons. A variation of this system uses Differential Global Positioning (DGPS) to determine its location.
- On Board Infrastructure Database. This data base includes information on the track structure (location and identity of beacons, location of switch points, track grade information by location, civil speed limit restrictions by location, location of interlocking stopping points, location of station stopping points, etc.).
- Train characteristic database. The propulsion and braking characteristics of the train. This information is necessary to calculate the safe braking distance of the train.
- Data channel link to a wayside zone controller (See Data Communication System(DCS))
- Basic Operation. The train calculates its position on the track and sends a data message with this information to the wayside zone controller. The wayside zone controller sends a message back to the train with information as to the train's limit of movement authority (the point on the track ahead that it has permission to proceed to). The train then does an on board calculation of the safe speed profile to the limit of authority point. This takes into account the civil limit restrictions over this path and the safe braking profile that must be respected in order to stop at the limit of authority point. As the train moves ahead it periodically (typically greater than once per second) transmits its new position to the wayside controller and receives updated information from the wayside as to its new limit of authority. If the limit of authority is based on the position of a train ahead, then moving block functionality is obtained.

Wayside Functions

- Zone Controller. The wayside zone controller manages the movement authorities within a contiguous track area. The zone controller receives position information from all trains with its control zone as well as information about train positions and route settings from adjacent zone controllers. The wayside zone controller also receives information about route alignments (point alignment) from an interlocking controller. This may also include information about track circuit occupancies and signal status if the system must operate with mixed mode traffic (trains that are equipped with full control capability and those that are not equipped). The zone controlled manages the information from all sources and then sends to each train a message indicating its most restrictive limit of authority as well as the points alignment on the route from its current location to the limit of authority point.
- Interlocking Controller. Some railroads require that this be an independent function. If this is not the case, then the interlocking functionality could be incorporated into the zone controller. The interlocking controller receives the position status of the points and controls the points position. The interlocking controller may receive a route alignment request from a centrally located operator (CTC) or it could receive route alignment requests from an approaching train via DCS message. The interlocking controller must have the position information of all trains within the interlocking (and it's approaches) in order to establish a route. This information may be obtained by track circuit status information or by the train position report, depending on the system.

Data Communication System

- Provide wireless communication capability between a train and the zone controller. This could standard radio data channel, spread spectrum radio, or a wave guide/leaky coax system. Bandwidth requirements depend on the number of trains within a zone that are using the same channel, the data message length, and the frequency with which messages are sent. The time interval between receipt of train location messages and the sending of movement authority messages directly affects the train spacing (headway) that can be achieved. Note that the zone controller that a train is communicating with changes as the train moves from zone to zone. The DCS system must be able to handle this transition.

- Provide communication capability between adjacent zone controllers. In order to provide smooth transition for a train moving from one zone controller to another, the train needs the limit of movement authority in the next zone ahead.
- Provide communication capability between zone controller and interlocking controller if they are separate units. In a distributed system, the area covered by a zone controller may not coincide with the area covered by the interlocking controller. An interlocking controller may be communicating with more than one zone controller.

Qu 6. I am particularly interested in the problem of controlling the lie of points from an approaching train rather than a central interlocking. I would be very interested to learn whether this can actually be done currently and what constraints would apply to doing it (if it has not been done, I would be interested in your opinion of what could be done)?

In the US we generally control interlockings by having local vital systems (relay or processor based) control each interlocking. Non-vital requests are sent to the interlockings but all the vital control is done locally. The systems have a central non-vital system responsible for determining what routes to select but the actual implementation must be considered safe by the local systems. In some systems (transit) route requests can be triggered from the approaching train but again the decision to implement is done by the local vital control system.

We have not done a system with direct control of the individual point machines or routes from the on-board system. We have discussed this with some railroads and believe it is possible. The major benefit appears to be in allowing work trains or trains picking up and dropping freight cars the ability to manage an interlocking area without involving the central operator or dispatcher. There may also be some advantages in areas where communications to the central office is difficult or expensive. We discussed a system that would give a train a limit of authority (LOA). This LOA would allow the train to move with in the defined track limits. The train would determine its movement authority limits based on the status of wayside equipment, the location of other trains in the area, and civil speed limits within the LOA. The train would then (using its location determination system and an onboard track map) decide when to call for switches and/or routes within its LOA. Once the train received conformation from the wayside local equipment that the switch was set and locked then the onboard system could move the train's movement authority to cross the switch. I think this approach is similar to the token block approach.

Qu 7. Are you aware of any projects / research towards the development of rail infrastructure that would not require moving points? (If you are, I would be particularly interested in details as this would be of major interest to both vehicle based control and moving block system capacity).

The only systems that I am aware of are people mover type systems (rubber tired), where the guidance mechanism steers the vehicle in the proper path. This usually uses a guide wheel/entrapment mechanism on board the train that engages a guide rail/beam on the wayside. Onboard deployment of a right or left side guide wheel causes the train to take a normal or reverse route through a diverging switch. A Maglev type vehicle could also use such an onboard method to determine the route taken through a switch. I am not familiar with the mechanisms used on these vehicles.

Qu 8. In order to determine the parameters that should be applied for realistic simulation and assessment of Moving Block systems and the potential for Vehicle Based Control, I would be interested to know the system delays that can typically be expected for:

a. **Trackside data processing delays** *(i.e. how long it takes an interlocking or equivalent control system to produce a new output following a change of its inputs). I think that there will be several groups of times for this, including:*
   - *delays that occur entirely within the processing element (i.e. the delay between receipt of information at the interlocking and a revised output)*
     Usually this takes between 0.5sec and 1.0 second depending on the interlocking equipment (relay or solid state)
   - *delays that occur in generation of the inputs (i.e. the delay in train detection equipment between the time a train occupies a new track section and the moment that the interlocking input changes)*
     Typically this is between 1.0 and 2.5 seconds depending on the type of track circuit that is used.

b. **Trackside equipment delays** *(i.e. the delay between a set of points being instructed by an interlocking to move and the moment that the interlocking receives a signal confirming they are locked and detected as requested)*
   This time varies greatly with the type of point machine used and the power source available. In the US, an electric motor powered point machine that has 110v power

operates in about 3 seconds. One that operates from 24 v may take 8 to 10 seconds to operate. An air operated point machine is faster, typically less than 2 seconds.

c. **Transmission delays between trackside equipment and trains** *(i.e. the delay between an interlocking determining that data should be transmitted to a train and it actually being received). There will be several scenarios for this transmission:*
   - *by free radio transmission*
   - *by guided radio transmission*
   - *by spread spectrum transmission*
   - *by coded track circuits*
   - *by continuous conductor loops*

   There is a great deal of variation in these times within a technology. The factor most influential is the access method. These include polled, time slot assignment, and collision avoidance. A second factor is the number of individual message sources that must be accommodated within a channel (number of trains, wayside locations, etc). A third factor is the message length. And the fourth factor is the data rate.
   There is not too much difference in the first three methods above, assuming they all have similar bandwidths. With current systems the time is in the 0.5 second to 2 second range. The coded track circuits are low operating frequency narrow band systems. The amount of data that can be transmitted is limited. Typically these systems take between 2 to 4 seconds. Continuous conductor loops are somewhat higher frequency than track circuits. I would rate these systems in the 1 to 3 second range depending on the message length and the number of trains that can reside on the loops at the same time.

d. **Trainborne processing delays and reaction times** *(i.e. how long it takes from receipt of a message for the trainborne system to interpret it and change its output states in response)*
   This is highly dependent on the onboard processing equipment and the complexity of the functions it must perform. With current equipment, this is in the 200 millisecond to 1 second range.

e. **Transmission delays between trains and trackside equipment** (i.e. the delay between a train determining that data should be transmitted to the interlocking and it actually being received). There will be several scenarios for this transmission:
   - by free radio transmission, by guided radio transmission, by spread spectrum transmission, by coded track circuits, by continuous conductor loops

   These times are similar to the times stated in c. above since they are the usually the same channel in the reverse direction and use similar techniques.

f. **Transmission delays between trains in the same area** (i.e. the delay between a train determining that data should be transmitted to another train and it actually being received). There will be several scenarios for this transmission:
   - by free radio transmission, by guided radio transmission, by spread spectrum transmission, by coded track circuits, by continuous conductor loops

   We have no experience or knowledge of systems that directly communicate train to train. However information that goes indirectly from train to train can be estimated by summing the above factors.
   For example: Train A sends a message to wayside control unit X with information about its position. Wayside unit X receives these position messages from all trains in its zone of control and arranges an "occupancy map" of this zone. Wayside X then sends a message to train B (which is behind train A) that its limit of movement authority is the reported position of the rear of train A. The total delay would be the transmission time from train A to wayside X plus the processing time for wayside X plus the transmission time from wayside X to train B.

For all of the above points, I would also be interested to know of any differences that arise in system delays due to the choice of centralised or distributed control architectures
In our opinion, distributed architectures are generally faster than centralized architectures. Communication delays are usually longer with the centralized architecture because other data links are required – usually land line data networks or radio repeaters to carry information from the zone base station back to the centralized location. Also, the processing time for the centralized processor is longer because it has more information to process.
A geographical area that would have better performance with a centralized system than a distributed one would be a very complicated interlocking arrangement with multiple tracks, such as might be encountered at a terminal or station (an example would be Grand Central Station in

New York City). A centralized interlocking control would be superior to a distributed architecture because of the interrelationship of the various routes through the terminal area.

David McKeown TD BSc(Tech) CEng MIEE FIRSE MIAM FRSA (Director, Creative Engineering Solutions Ltd).

Questions Raised by D. Woodland on 20/04/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I have mainly worked in maintenance all of my life. I started off as a British Rail Management Trainee, with Telecoms as my speciality, which at the time was very different from signalling. I was then the PA to a board member. After that I was in the HQ Signalling Development team, where I was involved with things like the first pilot scheme of SSI at Leamington Spa (trying to break it) and C-APT (the control of the Advanced Passenger Train). After that I escaped board headquarters and went to a line management job in maintenance. Then successive maintenance and S & T jobs, up to Regional S & T Engineer. Before privatisation I worked on Network Southeast with Chris Thompson, just after the Clapham accident in 1988, so dealing with all the changes to standards and the problems and reorganisation after that. After that I helped set up a consultancy called Opal Engineering, which has now become a part of WS Atkins Rail, and now I am an independent consultant.

At first I resented the IRSE and signal engineers, because I thought that a wider view was important and did not see the black art, with nothing written down, as impressive. I became a signalling engineer because that was the only way to be appointed head of an S & T department. I therefore took the IRSE exams as a signal engineer (despite being a Telecom Engineer at the time) in order to prove that I could understand the signalling too. I have always been the kind to challenge things. I have also been more a manager than an engineer. I have always been interested in getting things done through people.

My experience as a maintenance person has always been accepting projects or not, and I have been fairly critical of the way that they have been done without thinking right from design to installation. For instance, I was the recipient of the last free-wired relay interlocking at Westbury, where you had large palaces of buildings and there was more information available to the technician than the signalman. You could see the aspects downstairs, but upstairs on the panel you could not – you just saw red or green. I became very unhappy then about the whole thinking of capacity, regulation and management of traffic. There was a patronising view, and there still is, that signal engineers would come in and set up a system. When it was working the operators could use it, but when it was not we would take all of our toys away again and leave the operators in a void. This lack of understanding of the drivers, signallers and operations planning viewpoints, even timetabling, needs to be overcome in the future.

Qu 2. In your recent presentation to the IRSE, you raised a number of questions that I would like to address within my PhD research. I would therefore appreciate knowing what answers you would give to them yourself. In particular:

A. You mentioned the subject of cross-modal contrasts. Could you identify any such contrasts that you consider particularly worthy of investigation, together with an explanation of why? (I would particularly appreciate details of any papers, etc. that you may be aware of in relation to this).

One comparison that I think would be particularly interesting for you to look at is the concept of slots for air traffic. The concept that as you can't simply stop an aircraft, you have to think about getting to your destination and landing. Therefore there is an awful lot of available theory. It may not always be terribly well done, but certainly in European skies where they are very busy, you don't take off before you have your slot. You may not have all of your airways sorted out, but the capacity is allocated even if the exact slot is not known. The landing slot is at least known to be available. If we had that on the West Coast Main Line we would never have traffic interference - because the pathing would be sorted out in such a way that you don't just look at a little bit (eg as we leave the station), but would be checking that the path is complete. That is you don't interfere with the path elsewhere unless you can understand that you can cross it without obstructing it. I think that is a very important comparison, which changes the way you look at the management of the network.

Any idea where I could find information on that?

I don't have it at my fingertips, although I can give you contacts within the air industry. Also, if you talk to people like Praxis, they work in the air traffic industry anyway. Similarly, a lot of people working on the West Coast control systems will have worked for the air control industry in the past.

There is interesting work to examine done by the Universities of Duisberg and Stuttgart, also the Dutch Highways – and have a look at Intelligent Transport Systems, IEE Road Transport Industry Group (start with the IEE Review March 2000). I think that another very interesting concept to consider would be that of road trains. If you look at some of these areas, I am looking at the management of groups of vehicles that are either becoming or not becoming attached to other groups of vehicles. There is an awful lot of work that has been done there that ought to be leading our thinking. After all, what is a train?
I think that Bristol University is doing a lot of work, or at least Bristol is the model for the project, for a personal mass transit system. Bristol has been chosen because it is an old typically London like place with lots of wrinkles and roads are not a grid system. They have been looking at models for a personal transport system having minimum vehicles but actually on demand. They have to make best use of the roads so that you can still have cars and busses but these new personal mass transit vehicles mix with it. So they are looking at speeds and at how to control them. They have now got to the point where they are trying to get funding to do this. The idea is a bit like trams, where they will interleave on the roads, and head off into the suburbs. This is more than just research. It is now a project proposal out for funding to use Bristol as the pilot scheme for the city transport system of the future.
One of the interesting things about the pursuit of things like personal rapid transit schemes is that they will have to address what is an acceptable risk for roads. At the moment that rigour for roads is missing. A road is just a bit of tarmac. Who is limiting the traffic on it? In some places there are traffic lights to stop you getting on to motorways, but based on what theory? So once you make road traffic disciplined rather than chaotic, you are actually increasing the comparison with more disciplined forms of transport such as railways. I think that will be a good influence.
In terms of cross-modal, I think those would be the two most useful areas of research. I think that the day of the mixed traffic railway is gone. We are going to end up with the S-Bahn, the metros, the mass transits, high speed passenger, LGV's, and maybe more than one type of freight railway. Then we could adjust our systems to the specific needs of the traffic. The usefulness of a mixed traffic railway is I think is disappearing. So, when you look at cross modal issues, try to think of not just one railway comparison. You need to consider the effects and opportunities for all of the types of railway that we could have in the future.

B. How would you define the capacity of a Railway system? Please state what you would consider to be appropriate units for your defined measure.

I am not going to give a tidy answer to this. This is something that academic research ought to be considering, because I think it is something that is not really understood at the moment. However, I will make a few comments.
I think that there is a difference between theoretical and useable capacity. You de-rate the available capacity for resilience of delivery. (See also my IRSE Paper!)
Also, the stress is another interesting thing. What is the capacity of a power line? It depends on whether you are prepared to let it glow red-hot and risk it burning out, or whether you are most concerned about commercial advantage - then you will probably think about losses and things like that. Again, what is the capacity of a pipe to contain gas or a fluid? Well, it depends on what pressure you are prepared to put on before the risk of the joints bursting becomes too high. So what is the capacity of a railway? It depend on the risk you are prepared to take that something will go wrong.
The approach to the capacity has to be ultimately goods carried. So it is not just trains per hour, it is freight tonnage or people. But, it shouldn't just be people on the route that the train takes. Customers want a through journey, so the capacity needs also to measure the delivered connection as well as the travel on the train. If the connection is not properly delivered, then you effectively have leakage. Passengers will either not travel if it happens too often, or if the delay is too long will divert to a taxi or some other mode to complete the journey. You need to consider the complete journey and experience.
At the moment, most of the post Hatfield speed restrictions have been lifted, but the timetable still allows for them. As a result, we have a lot of trains running early. They then have to stand and wait in platforms for several minutes. What is that doing to capacity? Because of the design of the stations, there are not lots of excess platforms to absorb the delays, and because of the design of the timetable, this is not just a temporary occurrence either. In the days when train punctuality was measured purely by time of arrival at final destination, the journey time from Reading to Paddington was completely different to that for Paddington to Reading. All of the recovery time was in the last part of the journey, to ensure that trains always ran 'on time'. So, when considering capacity, you need to consider how thick the perfect path is, and what the tolerances are that you are prepared to put up with. The capacity will be a complicated trade off. The arrogance of the signal engineers has forced people to think of capacity only in terms of trains per hour, because it is simple and because it can be thought about in terms of block sections, and because that is the way that we do it. It is a given that they have been prepared to take based upon assumptions like the maximum line speed, the driver reaction time and behaviours.
For existing railways, you will need to look differently. If you have a track circuit block railway, then you have a limitation on what can be achieved. Trains per hour will be the basic measure. Things like the length of a train will have a big effect on capacity for a fixed tph. However there may be limits to the

size of train that can operate within the fixed infrastructure. For instance, at Borough Market, some train lengths are literally within inches of where the block joints are in the track circuits. If you made the train any longer it would foul them, and then you would literally be unable to move any trains.
In general, you are going to have to look at a generic academic point of view, regardless of the type of signalling. Something like erlangs used in telecommunications is required – but a single metric is unlikely to be the best approach.

C. You raised the subject of vehicle based control. Are you aware of any particular research into this subject (either from the railway industry or other industries) that you would recommend I familiarise myself with?

I think that one of the sources would be Westinghouse. I know that they have some patents in there from way back. You may want to talk to Tony Howker, and Terry George as well. There has also been a lot of useful work done in Germany. Also the air industry. Arguably, if you went to see the coast guard and the people who observe what happens in the channel, that may be interesting too.

D. Of the long list of railway control related issues that you raised as requiring academic consideration, which would you consider to be the priority, and why?

If we consider the railway as a series of pipes and islands, at the moment we have quite thin line railways and stations are only a little bit thicker. If you are going to put vast numbers into high speed railways, the implication is that you have vast traffic and when that slows down that needs a lot of space, so the shape of the railway needs to be very different. The view of the problem that most people have is misguided. The problem is not the main line but understanding and allowing for and arranging and controlling and managing the other bits at slow speed. It is also the bits at slower speed that allow us to get where we want to be, but the cost will go up, so that is where the problem is. That leads us towards a topography that looks very different to today's railways.
Have you thought about speed limits? What is the speed limit of the motorway and on a railway line? I suggest that you find out where they come from. I think that you will find there isn't any formal reasoning. Go back to the idea of the pipe. If you take a motorway and say there is a speed limit that does not get observed, and there is no consequence of that, what is the effect of higher speeds on capacity? If you have a railway with a more disciplined approach (that you absolutely do not exceed speed limits) then that line speed is quite important. If you set too low a ceiling, you have unnecessarily inhibited the capacity of that line.
If you are setting that ceiling on the basis that you are not exactly sure what it should be and that trains may not observe them anyway, then you have lots of unsystematic thinking with people allowing margins in ways that are not thought about in terms of their interactions. I think that this is a very important area for further consideration. If you look at the state of the ride now on Railtrack infrastructure, arguably one of the reasons for a line speed is to stop things dropping off. So if you could control the alignment, cant, etc well enough, then you could afford to raise line speeds (assuming they would not be broken). If you could be sure of the real conditions, there could be a large payback available – but it would throw a lot of the threat back on to the calculation of what line speed should be in the first place.
There is not the adequate theory there at the moment. This is even more true of temporary limits, like 20mph. You could run without rails at all at those sorts of speeds. Because the people responsible do not understand what the safe speed would actually be, they just put on a very low speed, without any understanding of the effects that that will have up and down the pipe that forms their network.
If you consider the result in terms of having to stop trains and out of course running, what has actually been achieved by the introduction of that restriction? If we do not understand where the line speed came from in the first place, we can not possibly understand what has been achieved by a restriction to it. There are some very big questions there.

Qu 3. The terms 'Signalling' and 'train control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control system'?

Signalling has always been seen as train separation. It hasn't been about the delivery of a journey or a timetable, or anything like that. The development of signalling has been largely evolutionary, based on accidents and preventing human error. I see it as a very basic system, which is only available when it is fully working, and helps train separation.
Train control has a dual meaning. One is the control system for the train itself, and the other is the control of the train network. There is a problem there with the terminology that we use.
If you look at the way that the railway has developed any of its thinking, BR-ATO for example came around in the late seventies, early eighties. A parallel development was a thing called junction optimisation technique. In the same way that TPWS is related to ATP, junction optimisation technique was a spot application system to deal with the issue of mostly converging junctions where you have got a mixed traffic railway and should be giving priority to the express. It was a way to do that without needing the signalman to do it, by use of the head code to try and set up the route. It only considered the particular junction rather than the whole network. Based on the days when any of the systems were first

come first served, it tried to decide first of all the priority of the traffic, and then the out of course running - how long to hold the junction for a train that was running late.
The first thinking was about how to solve that problem, but it evolved into network thinking. So, you have the basic signalling to keep stuff apart, and now you are starting to introduce a bit of regulation (what to do about those junctions) and a bit of thinking about the train itself (how to automate a bit more of it – although still focused on what is in the cab). The first thinking was then to have some person onboard the train who pressed the go button (like the Victoria line, but on a main line railway). They tried to address all of the issues about what was the weather, what is the braking force, what is the weight of the train.
If I return to train control, I have this problem as to whether train control is the control of train systems or the control of trains. I don't know that these terms are very well separated. I tend to think about ATO. If you talk to a metro engineer, train control would already include regulation to them, whereas main line engineers would probably be more inclined to interpret ATC as meaning ATP.

Qu 4. Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

The problem of confusion becomes worse when you start to talk about automatic train supervision and automatic train regulation. I think that regulation is easier, because it has always been used to refer to spacing trains out. It has nearly always been a matter of how long you hold a train for before you let it go, with the assumption that you would never let it go early. ATR is I think therefore fairly straightforward.
ATS suffers from the myopic thing. Is it supervision of the onboard systems, fault diagnostics and that sort of thing, or is it network supervision of trains and their operation, including regulation again.
I think that you really need to abandon the traditional terminology, in order not to get bogged down in the different meanings that are attached to each of them by different people. The problem is also that the initials can stand for different things too.
The hierarchy that you come up with needs to be like the development of a railway. One engine in steam does not require any control. Signalling as well as giving separation and therefore the ability to move trains without colliding, is the delivery of capacity. So if you want to run ten trains a day from Manchester to London, you can run ten parallel tracks, or a system that allows you to run ten trains down the same track at different times. So it is time-interval multiplexing. So if you consider telecommunications, you could really do worse than get an understanding of erlangs and modern data packet switching, because it is not really very different to interleaving on a fast moving railway. If you were to adopt that sort of terminology, that might make it a lot more neutral and get away from the heritage of misunderstanding. So I would say that the hierarchy would start with the basic separation, but without worrying where they are in quite large blocks.
So if you were to think about digitisation of speech or sound, the number of levels of service and the frequency with which you sample the call is a bit like how big is a block within which you want to identify something down into. It is a bit like GPS - to what level of accuracy do you want to have your data? Obviously the problems are that the consequence is that you have to know more and record more and deal with more volume of data to manage to find out the issue. So this is really the thinking that makes me think that you should invert the architecture.
So you have the signalling, which typically did not do anything about the operating. It was the necessary but not sufficient condition for running the railway. Then you have 2 levels of automation. You have the automation of on-board activity, but also the automation of network operation. That is where these things get muddled and it is important for you to separate them.
The abandonment of individual point control system for signal boxes to centralised control of points I think is another area where the system is creaking - because the overhead of taking information back to the central point and the time taken now to get it back are actually the dominant problem. That really is what forces you to start thinking that you should make the individual vehicles sort their way through the jungle, don't try to control all of the animals in the jungle at the same time, because the propagation, processing and algorithms then become far less. All of those issues are quite interesting.
If you go back to the old system, for example the anti-pre-selection controls, why are they there? It is worth thinking through some of those issues, because in some ways pre-selection is a good thing. A lot of current automated systems are waiting to try and set a route according to the timetable. There are some quite subtle things in there.
That is the sort of hierarchy I would be looking for. So you have a basic railway that is safe, a basic railway that will run a bit better, you then have a railway that has fewer humans in the different roles, and then drawing out the control units. Where you are taking humans out is effectively where you are putting automation in. You may want to find good terms for identifiable shapes of that. Then you have to start

thinking that there are some very complex control loops. Why don't we have intelligent vehicles that could find their own way through. So with a really good planning system which understands resilience, train braking speeds, and all of those things, and then taking hands off and providing the information for the Vehicle to make its way through. So it says please set me a route. How far ahead? That is an interesting trade off. It is what aircraft do, which is why I think that slots would be a useful comparison. When do you hand over to airport control (radar / approach / visual) and when do you simply look after your own separation. Particularly in airways, all you need to know is where the aircraft ahead of you and the next one to the side or behind are. For railways you will have less degrees of freedom to worry about in doing that.

You need to separate proprietary ATP systems from theoretical ones. Just because a real system works in some way doesn't mean that is how it should work.

The other thing that I would come on to is the whole of system thinking. That includes passengers and stations. Have you ever stood in a terminal station when an announcement gives a platform alteration? Chaos ensues. What has that done to the passengers journeys. There was an interesting IRSE paper 'signalling the passenger'. I suggest you read that.

Then we also need to consider graceful degradation and fault finding, which is another layer of what I call supervision. If you understand supervision in the way it is used in telephony, that may be a better way of using the term. That is call supervision, the setting up of the call, but not the content of it. Modern switches gets involved in setting up analogue calls (eg PABXs) and then get out of the way.

Qu 5. I would welcome any views that you may have as to ways in which ATP could best be optimised to create the greatest benefit for UK railways in both the short and long term, and what ATP developments you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

I don't think that I am interested in ATP. (Enough work is being done already – but if it only creates an envelope, very expensive and may not be successful anyway, eg driver overrides.) ATO yes. That is where you start getting payback.

Qu 6. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

One of the big things that tend not to get looked at is assumptions, and the false sense of security that ATP engenders. A lot of operators genuinely believe that ATP will prevent all collisions, and it won't. There is a whole range of concerns that come out of that.

If ATP won't prevent all collisions, what would you need to prevent them all, and shouldn't we be looking at that as ATP instead? How can we talk about trains being protected if they can still collide? Again, I think that you need to abandon the traditional terms and when you look at hierarchies, consider what is actually delivered. It is actually a partial train protection. It is a throwback to the old signal engineers attitude that I will provide a system until it goes wrong, and then take it all away. That is just not good enough. That is not a system!

There are too many administrations people, contractors and factories still thinking completely wrong headedly. If you really think that we will ever be able to operate systems with a half-second headway at full speed operation without some very different thinking then you are mad. And think about other people who are trying to develop control systems and theory for automated highways and personal rapid transit. Do you think that there will be a role left for signal engineers when they have done it? No. Those systems will be just as capable of operating a railway.

Qu 7. As an additional point to this, how do you think the trade off between safety and operational flexibility can best be managed?

You need to avoid words like new risk or introducing risk. Those sort of terms do not improve the chances of that sort of development being adopted. If you talk in terms of transport and assessing the risk, decrease of, the changing balance of, and other more neutral terminology, you are actually liable to be able to make people a bit more open minded about the threat. If you talk about bringing in new risk, before you can really explain what you mean people will put the shutters down.

Qu 8. What information do you think:

A. Is required by a driver in order to perform the tasks required of him

What are the drivers tasks? I think that you need to look at the system, which includes a human element, and describe what that element is. You may find that the IRSE Signalling Philosophy Review working group 2 report is quite helpful in doing that. Thinking in terms of what you would need to replace the driver, that would be distance and time to the next update of distance and time (the frequency of the cycle). Then within that cycle, what is the line speed (the ceiling / authorised speed, which may be to do with the location of a train ahead, or the condition of the infrastructure)? You then need to know what is the authorised distance to go, and probably the distance to any obstacle (which may be different if you are not extending the movement authority as far as the next train for reasons of operational flexibility. So, that is probably the basic elements needed to replace the driver. The driver does not therefore need to know the route (which is why I prefer speed signalling).

B. Would be beneficial to a driver to assist him in the performance of his tasks

What is beneficial to a driver in how he does his task, whether he should be there and what he can add is another whole ball game. I immediately think of the pilot. The consequence of too much autopilot is that he is not then capable of taking control in an emergency. If we talk about ergonomics, if there is anything less ergonomic than most rolling stock cabs, I would like to see it! That doesn't go for some modern ones. If we then think about trackside presentation and arrangements and how poor that is. Who cuts back the vegetation and who thinks about the effects of multi-story buildings day or night /lit)?

Then what is the driver's goal? It is perfect on-time arrival, which will mostly be about slowing down. The irony is that you can not ever do better, because you can not improve on the most efficient timing. You probably want a path that gets you there at X-2, but like the Swiss railways you slow down and come in a few minutes later than you could have got there, because that is reducing your system stresses, and actually people don't like being early.

Although what is on time? People might actually like to be 10 seconds early, so that they can be on the platform at the arrival time – or is it leaving the platform that makes arrival? Similarly, should we depart at the advertised time if the doors are closed half a minute earlier and the platform is closed a minute before that? Shouldn't the advertised time be when the access closes, even if that means the train leaves 2 minutes later? Otherwise it is like the old saying that the railway would run a lot better without passengers!

So to help with achieving this task, the driver almost needs real time graphing that some continental systems use. The real time generation of train diagrams based on feedback from the signalling system. The signaller can then advise the driver by radio to adjust speed according to the projection of where trains will conflict. A bit like junction optimisation in advance - like air traffic control does it. Slow down because there isn't any space for you. If the driver can get you there exactly on time and use less fuel in the process, then he has done well. So you then get into things like coasting. So I would present some information related to that sort of thing to him.

C. Would be detrimental to a drivers ability to perform the tasks asked of him

Well, anything other than the above really!

An interesting comparison would be lifts. Is a lift a train? Do they have a driver. Really the Victoria Line is just a lift. The driver presses the door open and close button and off it goes. The automation has been put in, but you have still kept the driver.

You haven't really asked about systems, but until we start thinking about the whole system, that includes the passengers and not just signal engineers' interests, you can't really answer and look at a lot of the things that you want to look at.

Interview by e-mail with:

Mervyn Parvard , BSc(Eng), MIEE, MIRSE (GE Transportation Systems Ltd).

Questions Raised by D. Woodland on 05/03/2003. Responses given by M Parvard on 05/03/2003.

Qu 1. I was wondering if you knew the frequency with which trains transmit their location signal in the GE Transportation Systems moving block system implemented on BART?

The location information is transmitted from the train to the wayside computer every 0.5 seconds.

Aditional Questions Raised by D. Woodland on 25/03/2003. Responses given by M Parvard on 05/03/2003.

Qu 2. I am interested in knowing what size safety margin is required for moving block operation, but have found no mention of this in the papers about GE's AATC that I have. If you could give me any insight into the GE view on what is required (from BART or other projects that have been implemented), it would be a great help to me?

You sure know how to get to the difficult issues! This is not a simple or short answer, and it is my current understanding of what is implemented on the BART system.

AATC-fitted trains are continuously sending back range reports. From these, the wayside Station Computer calculates the train's location, using some very clever algorithms. The accuracy of location for any fitted train is guaranteed to be better than +-15ft. [A train's location is considered to be totally invalid if it exceeds +-60ft.]

Depending on site conditions (e.g. tunnel or open air), the actual accuracy of train location varies. In open air (or in tunnels with smooth walls and good clearance between train and walls), accuracy can be +-5ft.

In tight tunnels, or tunnels with cast iron segments which cause multipath reflections, location accuracy worsens to typically +-10ft.
Since the speed and location of each train is 'continuously' reported every 0.5secs, the way the safety margin is implemented in AATC is to use a simple multiplier on the location accuracy. This multiplier (last I heard) is 8 for BART. Thus, irrespective of the speed of the train, if the location accuracy is 10ft, then the safety buffer is extended to 80ft out from the front and from the rear of the train. The same rule applies to the train in rear, so, in theory, the closest the trains can get to each other would be 160ft. This safety buffer can obviously therefore increase or decrease as the location accuracy is recalculated every 0.5 secs. However, this is very simplified. For example if the train is detected to have entered braking mode and the Station Computer verifies that the train is slowing down, this margin is reduced by another algorithm, such that a train can get within 20ft of an obstacle such as a red signal or a set of points not detected. Hope this is clear enough for your purposes.

Qu 3. Your explanation suggests that the trackside unit determining train locations, and subsequently movement authorities, must have an idea of the accuracy of train position data. How does it determine this?

Is it derived from the fact that the trains determine the distance from each of their radio units to 3 or 4 wayside units in their vicinity and, knowing the exact location of those wayside units and the topography of the line, the zone controller can determine a location based on any combination of 2 distances - any discrepency registering as an error, or is it more complex than that?

Your theory is basically correct. Each wayside radio's position is precisely known, all wayside radios operate off a synchronised clock signal, and therefore all the responses are averaged out in a clever way to reach the resultant accuracy figure. Remember, all Vehicle radios in general are receiving from a minimum of one Wayside radio, but, probably, on average, up to 2 Wayside radios in the forward direction, and 2 in the rearward direction. So they can correlate against multiple fixed wayside reference sites. I believe there is also some line specific data taken into account e.g. where there are curves in the layout in surface areas, the radio signals will travel in straight lines, bypassing the curvature of the track, so some compensation for this has to be built in. In tunnels, this is not such a problem.

Interview by e-mail with:

Gilles Poitrasson-Riviere (Head of Solutions, Alstom Transport Information Solutions).

Questions Raised by D. Woodland on 11/05/2001. Responses given by G Poitrasson-Riviere on 14/05/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

22nd of September 1984 I started to work on SACEM project. SACEM is an ATC system (ATP + Cabsignalling + ATO)
1987 responsible for the development of Trackside application software of MAGGALY (Driverless Moving Block ATC)
1989 Responsible for Application software on ATP and ATO in ALSTOM SIF.
1993 Responsible for the developments of MTRC ATCR project (ATP + Cab signalling +ATO)
1996 Responsible for the R&D project : CBTC (Moving Block ATC)

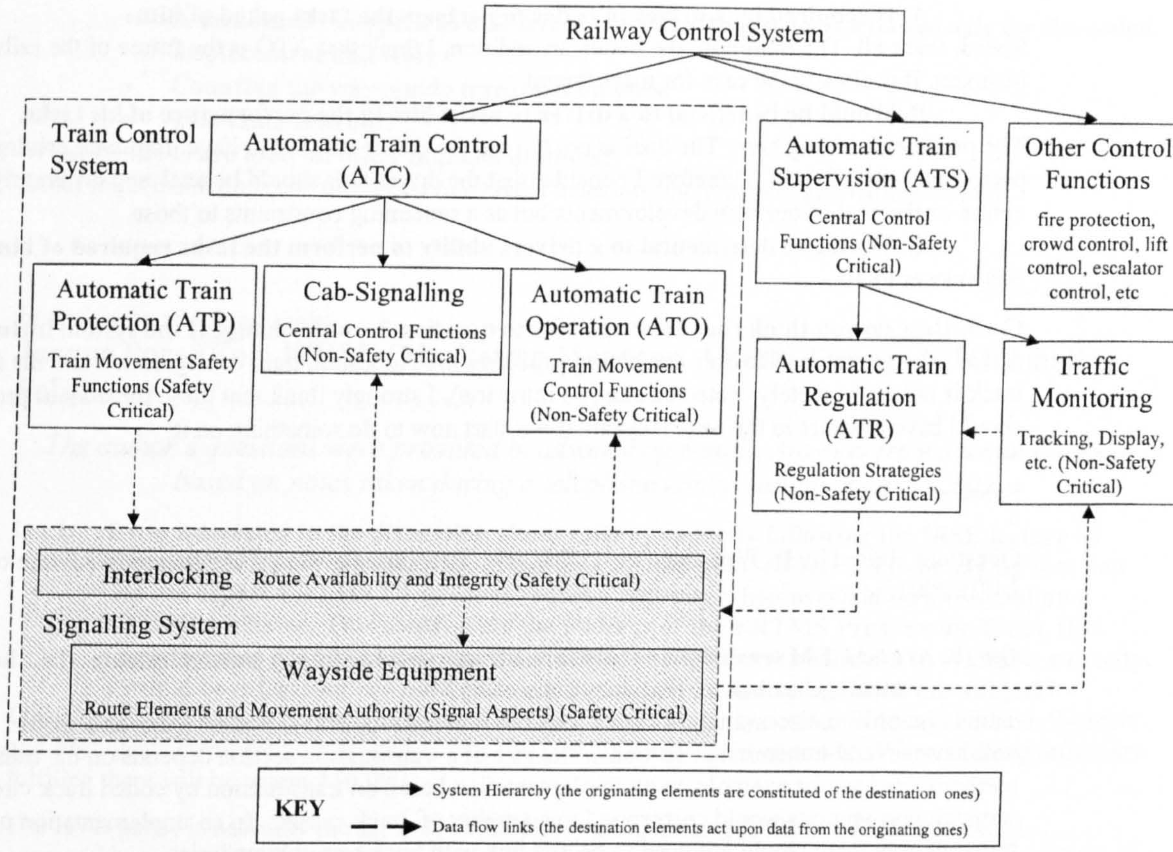
1998 WCML TCS Design Authority.

Qu 2. The terms 'Signalling', 'train control' and 'railway control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system?

Roughly:
Signalling = Train detection + Point machines + Level crossings equipment + Interlocking + Signals
Train Control = ATP + ATO + Cab Signal + Signalling
Railway control = Train control + ATS + additional functions

Qu 3. Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

Regulation is a function/part of the supervision which includes tracking, display, etc...



Qu 4. How would you define the capacity of a Railway system (and what would you consider to be appropriate units for your defined measure)?

Transportation capacity (how many people per hour), Journey time and availability
These two should include all the other parameters (speed, headway, train capacity, etc...)

Qu 5. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

Operational impact (is the customer ready to change ?)
History (how complicated is the current system ? How many specific situations ? What is the weight of history in it ?)
Requirements (what are the real requirements ? These are very rarely known. According to me they should be at the level or derived from the answer to question 4.
ATP requires odometry. The slip/slide phenomenon is very difficult to manage correctly. Globally Adhesion is difficult.
Configuration is also not an easy subject. Specifically for some parameters which are very difficult to validate (gradients).

Qu 6. I would welcome any views that you may have as to ways in which Automatic Train Control systems could best be optimised to create the greatest benefit for UK railways in both the short and long term?

The first step is to understand the links between the business requirements (new paths, question 4 answer) and the ATP solution. At the end of the day I am under the impression that it is very difficult to "justify" the investment because of this lack of link.
The structure of the UK railway business is also very strange (track to Railtrack, ROSCOs, TOCS). There is no one centre of decision.
After that the rest of the implementation of an ATP is purely technical and therefore simple.

Qu 7. The introduction of higher speed operation, higher capacities or changes to track layouts may all offer operational benefits, will influence the commercial viability of projects and will also affect safety. How do you think the trade off between safety and operational flexibility can best be managed?

The first step is to understand what is required. I am not convinced that it is the case today. Technical people take business decisions based on their experience and reaction to accidents, not on sufficient and correct studies. As these studies have not been carried out, I don't know what should be the trade off.

Qu 8. What information do you think:

A. Is required by a driver in order to perform the tasks asked of him

Speed, that's all. The minimum the better. In addition, I think that ATO is the future of the railway business. It is already the case for mass transit.

B. Would be beneficial to a driver to assist him in the performance of his tasks.

The problem is wrong here. The goal is not to ease the driver task but to have improved driving performanceof the train. Therefore I consider that the driver task should be analysed differently, not as the centre or the goal of our new developments but as a remaining constraints to those.

C. Would be detrimental to a drivers ability to perform the tasks required of him

See answer to 8B.

Qu 9. How do you think that your answers to question 8 would change if the system included ATO?

ATO is the next step. The only remaining problems are Adhesion, train configuration and the fact that the track is not a completely protected site (no intrusion). I strongly think that these three main problems should have answers in the next 10 years if we start now to do something on it.

Subsequent Interview:

Questions Raised by D. Woodland on 11/05/2001. Responses given by G Poitrasson-Riviere on 14/05/2001.

Qu 10. Are SACEM trackside to trainborne transmissions totally indescriminate (i.e. the same message sent to all trains in the sector)?

Transmission is indiscriminate within a transmission section. This is not the same as a sector. Each sector will have several transmission sections. The size of a transmission section depends on the transmission methods used so, for example, in an implementation based on transmission by coded track circuits the transmission sections would correspond to a number of track circuits. In an implementation using a radio transmission, there would not need to be any link with track circuit boundaries.

Qu 11. I understand that SACEM trackside processing can be performed completely independently of rolling stock characteristics. Can you confirm this?

That is basically true, but there is one exception. The trainborne computer has a limited processing capacity. In order to avoid system delays, some pre-processing of energy calculations is carried out trackside and the results for each train type using the line are included in transmissions. This could now be easily resolved by use of more modern computer with faster processing times. However, at the moment it does mean that changes in train performance would require some limited trackside data changes.

Qu 12. On the Paris RER Line A implementation, what mode do trains leave the depot in?

They would leave in 'CML', which is the French for 'free manual operation'.

I have seen this referred to as 'Standby'

It is not exactly that, but such a phrase could be used in translation.

Qu 13. Is any of the Paris RER Line A SACEM transmission equipment fail safe?

No, except that it is guaranteed that there will be no cross talk between beacons.

Qu 14. Are there any differences in functionality worthy of note between the Paris RER Line A and Hong Kong MTRC implementations?

The Hong Kong MTRC implementation provides ATO. It also has no Automatic signals (although controlled signals are retained in junction areas).

Something else worth noting about the Hong Kong MTRC implementation is that it can detect track circuit and transmission system failure and adapt its configuration in order to allow continued operation (albeit at a reduced capacity).

Qu 15. Are there any differences in functionality worthy of note between the Paris RER Line A and Singapore MRT implementations?

The Singapore MRT implementation provides ATO and has no signals.

Qu 16. I have determined form published literature that the Singapore MRT implementation determines train position by a combination of :

- **A coded movement detector coupled to the trains axle measuring displacement and direction;**

- **A tachometer coupled to a different train axle that detects a locked axle for the coded displacement detector;**
- **Counting the waveguide transmission slots;**
- **Trackside beacons.**

What methods are used on other implementations?
Other implementations use the coded movement detectors, tachometers and balises as well.

Interview by telephone with:

Jacque Pore IEng, FIRSE (Marketing Manager, Alstom Transport Information Solutions).

The author's questions were provided in advance by e-mail. Answers from Jacque Pore are based on notes taken during a telephone conversation on 16/10/2002.

Qu 1. I was interested in the discussion about railway capacity following the IRSE lecture on Eurobalises last Wednesday. Both the speaker and yourself seemed to be saying that you did not expect ERTMS L1 to cause reduced capacity when overlaid on a conventional signalling scheme. This contradicts the findings of the ERTMS Programme Team, that implementing ERTMS L1 in the UK will lead to significant reduction in available capacity. I would be interested to know more about your views and opinions on this subject?
I base my opinions on experience with KVB in France and similar systems in Norway, Finland, Sweden and Portugual. There are now around 120 to 130,000 KVB balises in operation. When existing orders are fulfilled there will be nearer 150,000.

The KVB balise is basically the same as an ERTMS balise, making KVB and ERTMS Level 1 comparable systems. KVB has also been installed on busy lines around Paris and Lyon, comparable to the lines around London.

Experience in France has shown that the installation of KVB did not have any noticeable effect on capacity when compared to the conventional signalling used previously. The discussion last week noted this.

There was also another point underlying the discussion that may be worth clarifying. Per Lundberg noted that ATP increases safety and that to gain equivalent safety without ATP you would need to extend braking distances and overlaps. This would reduce the capacity that can be offered without ATP. Generally, the approach to safety of traffic in the UK is not very good, because there is no ATP system and the operation is not carried out properly. Drivers are allowed to exceed permitted speeds by some margin when controlling their train to the line speed and trying to keep to timetable or make up time. This is not the case in France and elsewhere in Europe. If the line speed is 100km/h in France and the driver is found to have travelled at 101km/h, he is fined. If he does it several times, he will be removed from driving. In order to enforce this, all trains carry data loggers that monitor speed. 10% of these are systematically analysed to monitor performance and detect overspeeding. When ATP was introduced, the brake intervention was set to 5km/h above the permitted speed. However, drivers do not allow their trains to go that fast, so interventions do not occur in normal situations. Since UK driving practice is different from French practice, it is possible that an overlay spot based ATP system would require modified driver behaviour, which may affect capacity – but is necessary to ensure safety.

Does KVB require a lot of infill to achieve no impact on operation?
Not a lot. On plain line sections, infill offers no real benefit and is not used in KVB. If you used infill to allow a fast train to close up on a slow train ahead of it on plain line, you would just end up with the fast train accelerating on receipt of infill clearance, getting too close to the train ahead and being forced to slow down again (since it can not overtake), in a sort of sine wave effect on speed. This is inefficient and uncomfortable for passengers. The use of infill in that situation is both pointless and best avoided.

Infill is needed on approach to junctions and stations with multiple platforms. If the train ahead changes line and the route for the following train suddenly becomes free, signals can change from red or yellow to green. In that situation, it is useful for capacity to let the following train know about it. In that case, one infill balise about 300m before the signal is usually enough. On long block sections (such as the 1500m sections in France) a second balise 500m before the signal can also offer further improvements. There is no need for more than that.

In all, about 10% of KVB signals in France are fitted with infill.

During a previous IRSE lecture about the BR-ATP experience on the Great West Main Line and Chiltern line, it was noted that 'the introduction of an intermittent ATP scheme had an impact upon the performance of both routes' (Wright et al, 2002, p5). It was further noted that the effect on GWML (which initially had only balise based transmission, without infill) was 'considerably more pronounced'. In order to reduce this effect, infill balises and loops were fitted. In discussions that I have subsequently held with that author, he stated that the effects were still felt with the loop in-fill that now exists. This appears to be contradictory to your experience with KVB.

I do not think that can be right. It was certainly not our experience with KVB. I would be interested to know more about the scenarios that they experience problems with. I would not be surprised if it was more of a problem with operating practices than the ATP system. They would probably actually have the same delays even without the ATP.

Do you know what operational margins are added to the technical headway in developing French timetables?
I do not know that. You should try talking to someone with particular experience of SNCF operations. You may find it useful to talk to Moens Gilbert who works for the Channel Tunnel Rail Link about that.

The reason I ask is that an overlay system introduces processing delays that must lead to a reduction in technically achievable capacity. However, if the operating margins placed around the technical headway are sufficient, it is possible that you would not notice that reduction in the actual planned / used capacity. It is possible that this is the key to the different experience, since the UIC recommend that only 75% of available capacity should be used, whilst in the UK it is common to use nearer 90%.
You are right. I agree that on paper the capacity reduces when you fit an overlay ATP system. However, in practical operation it has had no impact on French KVB lines. I can not comment on the difference in operational margins between France and the UK.

Qu2. Was KVB purely an overlay on existing signalling, or did the underlying signalling have to be adapted in order to implement it without capacity impact?

Some re-signalling was carried out, but the majority of installation was as an overlay with no changes to the existing signalling layouts. Rail joints were not moved; overlaps remained the same lengths, etc.

What overlap lengths are used in France
Typically, in the region of 100m.

Interview by e-mail with:

Steve Rodgers B Eng, AMIEE (Senior Consultant, Booz-Allen & Hamilton Ltd).

Questions Raised by D. Woodland on 05/09/2001. Responses given by S Rodgers on 05/09/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?
From September 1993 to November 1995, I was a Graduate Trainee with the Engineering Directorate of London Underground Ltd. For the last six months of the scheme I was Site Engineer for the Central Line Project, responsible for the final design details and approval of new electro-hydraulic point machines used at 14 sites on the Central line. Also for the installation of these machines, associated cabling and Signalling Equipment Room modifications.
From 1995 to October 1996, I was the Assistant Resident Engineer, ATP System, on the Central Line Project. I was responsible for:

- organising test trains and associated resources, to allow test, commissioning and confidence-building train running to take place throughout the Central line;
- conducting the trackside and trainborne site testing, both with the signals contractor and independently, to verify operation in specification;
- collation and analysis of trackside and trainborne systems reliability information, identification of trends and recommendations for areas of improvements.
- audit and faultfinding of site installations before final commissioning activities.

Then from November 1996 to December 1997 I was the Automatic Train Protection (ATP) System Engineer for the Central Line Project. My responsibilities included the client activities associated with the testing, commissioning and reliability improvement processes for the ATP system provided as part of the Central line re-signalling. This included the production of weekly reliability analysis reports, site work as required and the conduction of the trainborne testing to verify the integrity of trackside systems. I also

completed the checking of WESTRACE electronic interlocking site specific 'ladder logic' data against signalling control tables prior to the re-signalling of the Woodford area in August 1997.
In January 1998, the Central Line Project became a part of Infraco BCV Ltd, and I was appointed as the Automatic Train Control (ATC) Project Engineer, the technical authority for the overall functionality, software, hardware and site-specific data of the Central line ATC system.
I left LUL in March 2001 to take up my current post as a Senior Consultant with Booz·Allen & Hamilton's Transportation Team, where he are working on various rail projects.

There are a few things that I would be particularly interested to know about the Central Line ATP system:

Qu 2. Are there any limitations that you are aware of in the Central Line ATP system
No particular limitations.
There are some reliability problems - mostly due to lack of / poor maintenance in LUL/BCV.
Drivers don't seem to mind it - bear in mind that the speeds the ATP will allow are often much higher than the train can achieve because of the closely spaced stations.

Qu 3. I seem to recall that there were problems with the speed code generators being used for a large area and on both roads. Can you confirm this / provide any clarification on the problem?
There were some reliability problems with code generators – but they now seem to be ok (some modifications were done). The problem you mention is about only having one generator for each frequency at each site (no backup) so if it falls over the whole site is affected (both directions).

Qu 4. I know that there were also interface problems with the trains brake system, but am not sure whether it was purely an ATO problem or affected the ATP as well?. Do you have any details about this?
There were no ATP problems here (but lots of ATO ones!)

Qu 5. I am also interested in more detail of what indications (visual or audible) are given to the driver / when. i.e. if I remember correctly he is given a target speed display in the form of a horizontal and another bar showing actual speed? In addition to this, does he get a warning of any kind if an overspeed is predicted?
Correct, the driver is just given a Target Speed display (integrated into the speedo unit). An audible warning sounds when the TS changes (two noises one for upwards change, one for down). There is no overspeed warning at all.

Qu 6. Is the trainborne system based on 2 out of 3, 2 out of 2, or is it just a single processor?
The trainborne ATP controller has two halves 1) Safety System. 2) Non-vital system. Either can apply the emergency brake. Both monitor the code from track and the train speed. They share some components outside the ATP Controller box itself (tacho, code antennas).

Qu 7. Roughly what frequency range do the coded track circuits operate in?

Carrier frequencies are 4080Hz to 6000Hz
Modulation frequencies are 28Hz to 80 Hz
Spot loop frequencies are 6240Hz to 6840Hz

Qu 8. Does the ATP use the target speed code in any way, or is that only used for an indication to the driver?
No, ATP only passes on the Target Speed to the driver & the ATO.

Qu 9. What happens following a brake application (when can it subsequently be released)?
The emergency brake is applied whenever train speed > Maximum Safe Speed. The ATP has no control of the service brake. The emergency brake will release as soon as the train speed < MSS, i.e. train is not brought to rest unless MSS = 0.

Tony Rowbotham BSc(Eng), CEng, FIRSE, MIEE. (Alstom Transport Information Solutions).

Questions Raised by D. Woodland on 30/04/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?
I have been working in the industry for nearly 40 years. I started a bit of testing on Nuneaton and Watford Junction. The first real engineering I did was designing location cases for Cheadle Hume to Grange Junction. In those days everyone used to start on location cases and progress up to relay rooms.

I did quite a bit on Trent and Saltly, which were central Reed systems. Then the first really big job was Dartford, when the guy who was due to design the relay room decided to leave and I was given the job. I knew nothing about geographical interlockings at all, and had to read up on them in the manual on holiday in Austria!

I also worked on the original Paisley job in Scotland, from the outskirts of Glasgow through Paisley and on to Gourock and Wemyss Bay. That was also a Geographical interlocking. At the time they would not trust having Gourock which at the time was a fairly major terminal station with three platforms and a number of sidings, on a remote control system, so it had to remain as mechanical locking (with colour lights) with a new lever frame.

After that I spent 8 years from 1971 working on the Kings Cross resignalling. It mainly took so long because of the Civil Engineering works re-aligning Hollaway station and remove some of the curves at Hatfield. I was responsible mainly for the indoor part including the interlockings, which was geographical. It also had a very complicated remote control system hierarchy. For instance, the system for Welwyn Garden City also included all the functions for Hatfield. The functions for Hatfield were then separated out at Welwyn and taken back to Hatfield in a separate system. That meant that the local control panel at Welwyn could also control Hatfield. Hatfield also had its own control panel, which meant that it could be controlled from Kings Cross, Welwyn or Hatfield! Lots of non-vital remote control systems with thousands of changeover contacts to select what controlled what.

That was my last real job! I then worked with Jim Waller in the Technical Directorate for years, working on standards, documentation and training. I have been running an 'Introduction to Signalling Course' within Alstom for about 10 years now. I have lost count of the number of times that I have done it! The course material has developed over time, although it hasn't altered much lately. I really should add TPWS, but it does mention ATP.

{Tony was a major contributor to, and organiser of, the IRSE 'Introduction to Signalling' text book.}

Qu 2. The terms 'Signalling' and 'train control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system?

I would argue that Signalling does not control the train at all. The driver does that. Signalling, if the driver responds correctly to the information given to him by the signalling system, ensures the safe environment for the public.

Train Control does control the train! So, when you have ATP as well, it is train control (because it does control the train at least when something is going wrong). Train Control is signalling and ATP (+ATO if applicable).

Railway Control Systems I am not so sure about. I think it would be controlling the railway in a wider sense than controlling the trains. It is trying to run the railway rather than just controlling the trains themselves. It would therefore include more of the ATS type of thing. So, Railway Control is Train control + ATS/ARS etc

Qu.3 Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

I would argue that Automatic Train Supervision is the basic setting of routes from the timetable.

I would argue that Automatic Train Regulation is a better edition of that, where the system can try to get the trains running on time if they get out of course. Regulating the timetable. That was what happened on Docklands, although it was called an ATS system there. It could tell the trains to use a faster profile if they were running late, in an attempt to help the service get back to the timetable. On Docklands the regulation function was fairly easy, just switching between two driving profiles. On other systems it may not be so easy.

What about when things go seriously wrong and trains need to be terminated, turned back or diverted?

That is not regulation. It is a different thing. We were talking about doing that on SIGNET. It was a sort of IECC system, but broken down into modules for flexibility. So, you had a module for route setting, one for ARS, one for talking to the relay room and another for talking to the panel. There was another module, the Forecast Module, that was going to be put in to predict conflicts in the future. So it would look at how trains were running and predict how they would look in, say, a quarter of an hours time, to see if any conflicts were arising and then try to do something about it before hand. The system was put in Delhi and Moscow (without the Forecast Module), but then dropped from more development when the Company was in a bit of trouble. For Docklands we were forced to use a GRS (now SIU) system.

Qu 4. How would you define the capacity of a Railway system? (please include what you would consider to be appropriate units for your defined measure of capacity).

I would use the number of trains per hour or per day. You can't talk about people because that involves the size of the train, so you could only do that if you had standard trains that were all the same. Even then you could have trains joined together in the rush hour, and the question of loading. So I would argue that capacity is trains per hour or per day.
Not so sensible is the time between trains. The system can be designed for 5 minutes between trains, but that doesn't necessarily mean that you will run trains 5 minutes apart all through the day.

The problem with tph is that it only applies to a given point on the network and not to the whole network service.

If you want to consider the whole network, you really have to start counting people – and that is not so easy. They would not be spaced equally during the day or over the network. So I think the easy answer is trains per day.
I think it depends a lot on what sort of railway you are talking about. On something like Docklands (where all trains are roughly the same and travelling at the same sort of speeds) it is probably much easier to measure capacity than on the West Coast Main Line (where there is a great mixture).

Qu 5. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?
I think that the main thing is how easy it will be. From that point of view, one of the biggest questions is can you fit the equipment onto the trains to start with?! If it is going to cost millions to even get the equipment onto the stock in the first place, is it worth even thinking about? They have done a lot of work on this for the West Coast. We will probably end up with a lot of stock running on the line not fitted. There was a specification out today, only guidance, saying that all new rolling stock should be designed to have the space and outlining what space is required. That may solve the problem in the future, but retrospectively there is still a big problem.
If a train already has Doppler radar for slip/slide protection, can we use the same unit for both functions? If so, that would help a lot. I am doing an exercise looking at these sort of things for the West Coast at the moment.
There will be big problems in this area. There is no point in worrying about the signalling side if you can't get the equipment on the stock in the first place. I think that is the main issue to consider.
If you consider the TCS MMI, I have a feeling that it is too big and complicated. If it could be made smaller, it would fit on to some of the stock a lot easier.
The you can look at the signalling and consider all of the different types that are likely to be found. You then need to decide whether it is practical to fit your ATP system on to it or not. For TCS we concluded that for colour light signals level 1 was generally OK, but that for level 2 you really needed SSI or another CBI type system. Then you can obtain the information needed from a connection to the systems general BUS, rather than needing to add any information into the system. At one time we were suggesting that you might need to put additional data into the SSI, which would mean you would need an extra dummy module just to act as an address to send the extra information to. On balance, they decided against that approach for TCS. For other CBIs there will generally be a BUS somewhere for the TCS to tap into, or there may be an output anyway designed for the interlocking to communicate with neighbouring interlockings. In that case, you could use that output to get the information out. So that is the other major factor to consider. If you had to actually start digging about in interlockings and getting information off of individual contacts, that becomes more of a problem. With a geographical interlocking for example, you may not have the contacts available to do it. Even if there are outputs provided, they may already have been used for some free-wiring anyway. So in that situation, you would literally have to look at every geographical unit separately to see if what you want is available.
So that is my argument there. Look at the rolling stock before the signalling. There is no point in wasting time with the signalling if the rolling stock won't support the system.

Qu 6. As an additional point to this, I would welcome any views that you may have as to ways in which ATP could best be optimised to create the greatest benefit for UK railways in both the short and long term?
Again, I would emphasise the fitment of rolling stock. Keep equipment to a minimum size and quantity. It appears at the moment that TCS requires 4 aerials on the roof (2 for speech and two for signalling). That seems a bit excessive. If you could optimise what needs to be put on the train it would seem a good idea.

Like the use of an inertial navigation system to replace most of the Doppler radar's and tacho systems required for TCS?

Yes, that sort of thing.
Then the other obvious thing has to be cost.

Qu 7. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?
If there are greater benefits, that would encourage fitment. So I would again argue cost. I have just read that the original ATP estimates of £14 million per life saved has now gone up to £100 million

Can you provide a reference for that?

I am fairly sure that it was in May Modern Railways which is now on circulation and fairly inaccessible. Also possibly in the strangely coloured Railway Intelligence publication, every couple of weeks and Roger Ford involved.

Qu 8. As an additional point to this, how do you think the trade off between safety and operational flexibility can best be managed?
Putting in ATP will increase safety, but at a great cost. I understand that the drivers don't like it because it reduces there flexibility quite a bit. What effect that has on the running of the railway and whether it affects headway I don't know. I also don't know how strong the feeling is or whether it is really true. The best way to get that question answered would be to talk to drivers who have used ATP.
Having said that, ATP may get rid of some driver quirks and force trains to be driven more or less the same, which might have some advantages form consistency.

Would ATP alone achieve that, or would you actually need ATO as well?

Yes, I would argue that once you have ATP on the West Coast, it would not cost very much to put ATO in. I would think that you could change the system so that the Movement Authorities include station stop information, add the traction control system, and not really have to change much else at all. You already provide distance to go, aspect and speed restriction information. Getting the train to stop at stations for the right amount of time could also be built in. That would help introduce regulation too, and reduce the dwell times. That would offer a lot of benefits and should be relatively simple to introduce. I would guess that the extra cost would be fairly small.

Qu 9. What information do you think:

A. Is required by a driver in order to perform the tasks asked of him
Not too much! A nice simple Speedo and what speed to aim at.

B. Would be beneficial to a driver to assist him in the performance of his tasks
Perhaps something to tell the driver how far it is to the next red aspect or speed restriction.
We actually have a patent out for giving the driver more regulation information on a London Underground type system by superimposing it on the VDUs they have in the platform to monitor the doors. That would help them to regulate their running to the timetable, giving messages such as the time to depart or 'you are running late'. The driver will be looking at the VDU anyway, and it wouldn't distract him with that information whist he was driving.

C. Would be detrimental to a drivers ability to perform the tasks required of him
I think that the TCS MMI is too big and complicated. There seems to be an awful lot of information on it. Are we giving the driver too much? If you give the driver a speed target, does he really even need to know how far ahead the next red signal is, let alone all of the gradient information. Just the target speed would be much more straightforward.
What about head up displays instead of the big MMI unit? Would that require less space and make fitment easier? It would allow the driver to look out of the window and see his MMI information at the same time. You could even envisage some form of head up display built in to a helmet.

Qu 10. How do you think that your answers to question 9 would change if the system included ATO?
You may not have a driver with ATO. If you did, there would be arguments for giving him more information than he would need as a conventional driver. Particularly if he only got to drive manually occasionally, you might need to give him coasting information, and other information to help him know how to drive manually. Under normal circumstances, it would be difficult to keep him occupied.

Qu 11. I am trying to develop an understanding of the historical development of signalling equipment, but have found no trace of anything about geographical interlockings. Could you please explain:

A. When they were first developed;
They were developed in the 1950s, introduced in the late 50s. The first ones were geographical only in the non-vital route selection functions, with a normal relay interlocking on the end of it. Then once the concept had been proved, they extended the system to include the vital interlocking functions.

B. Why they were developed;

At the start of the big 1960s modernisation Scheme, BR were concerned that there were not enough skilled people to design, install and test the interlockings that needed to be installed. So, they wanted a new method of interlocking that would be quicker in these areas.

C. How they worked, and their relative merits and problems;

The idea was that you had a package of logic including absolutely everything that was needed for a signal, another package of logic for a single end of points and one for double ended points. So, theoretically if you had a nice simple layout, you would lay out your units (or sets) on the relay rack in the same geographical layout (as far as you could) as the actual track layout {*Tony drew a sketch, as shown in Figure A1*}.

Then you connected each of the units together with nominally standard cables, in the same way as the equipment was connected together by the track on the ground. You then also had a nominally standard cable from each unit to the control panel, and from each unit to the outside equipment.

The units were connected together by an 8 core route setting cable. When you pushed the entrance button, a feed went out in one of the wires. It would then go through the points sets and according to point availability would pass along the route **only** to the sets of available exit signals. When you then pushed an exit button, you sent the feed back on one of the other wires. The points set would then see the feed come back, determine which way the points were required, and set the points that way. There was then a locking function, and locking proving circuits that would also go through this cable. There was then a separate 4 core cable, which dealt with clearing the aspects.

The way that the sets were organised, the entrance set would also have had the circuitry for the entrance signal overlap track within it. The points set would have had the circuitry for the A and B end track circuits within it, and the exit signal unit would have the berth track and overlap conditions, and the ECR or ECPR as well. So, the circuit would start at the exit set, feed back through all of the units collecting on the way all of the information needed for clearing the aspect. But, there were always holes left in for extra information, so the circuit would be brought out to terminals on the set. These would normally just be looped, but if you ever wanted to put anything special in as free wiring you could. That might be for an extra TPR, etc.

The idea was then that the units could be mode by the hundreds in the factory and tested on a routiner to eliminate most of the testing required for wiring on site. The drawing was fairly quick, since it was always based on a standard drawing. To install them, you just hang them on the rack, join the multi-cores up, connect the looping (and wire count it). The on-site testing was also very quick because you only needed to test each function once - the entrance function for routes from S1 would be the same regardless of whether the exit was at S3 or S7. You therefore only had to test it once, whereas a conventional free wired interlocking would have a separate circuit for each route, and each one would therefore need testing. The testing time was particularly impressive in a complicated area. The system also forced consistency in control tables. At a given exit, such as S7, the overlap circuitry for the routes from S1 and S5 is common to them both, so the tables must be the same for them both.

The trouble is, that a signal set to be totally generic has to have entrance functions, exit functions, route arm functions (how many?), counter route functions for moves the other way, and so on. The points sets had to have all of the functions that might be required for the points, overlap functions, facing overlaps, trailing overlap functions, etc. So really, if you strictly included everything that could be needed, the amount of redundancy (as in unused circuits) in each set would be immense. So, they were basically designed so that you could plug the relays in for functions that you wanted and leave them out for functions that you didn't want. However, that meant that you had to do a lot of looping on the set to loop out back contacts that were not there. They then needed to be wire counted and functionally tested.

The Westinghouse sets stayed like that. They were huge sets, very heavy, with over 30 relays in. What the old SGE/GEC did was split them down into sub-sets. So they had an entrance set, an exit set, a combined set for counter moves and route arms. That made for a lot less redundancy, but a lot more wiring (because having split them into sub-sets, you then had to wire them together – again using standard cables). Similarly for the points, the facing and trailing overlap circuitry was included in sub-sets. Despite the additional wiring, the system was still fairly quick to engineer. There was also a very old system by AEI, which was not made from sealed units, so you could make changes to the circuits themselves rather than needing looping.

Another problem was when you wanted to have a route arm on the signal. Because the circuit elements were common for all routes, you could not tell which one you had set back at the entrance set that held the route arm circuitry. A system was therefore devised, that relied upon the fact that you could only go through the process of setting one route at a time (which had also been the case for relay interlockings). There was another Set called a 'Common Control Set' that supervised all of the entrance and exit button selection. This monitored two standard cable rings. One went around all entrance sets, and the other around all exit sets. It could then make sure that one route could be set at a time, work out the class of the route, etc. In addition to that, there was another set of 4 rings called the route rings, which looped around the relay rack rather than the sets themselves, and was available to all entrance and exits. When you pressed an exit button, it energised one of the route rings. Whilst a route was being set, any entrance set UR's would also be connected to the rings. So as long as you were careful about how you assigned connections to the rings, they provided a means of telling which route had been set, and then lighting the right route arm. See Fig A2 for better explanation!

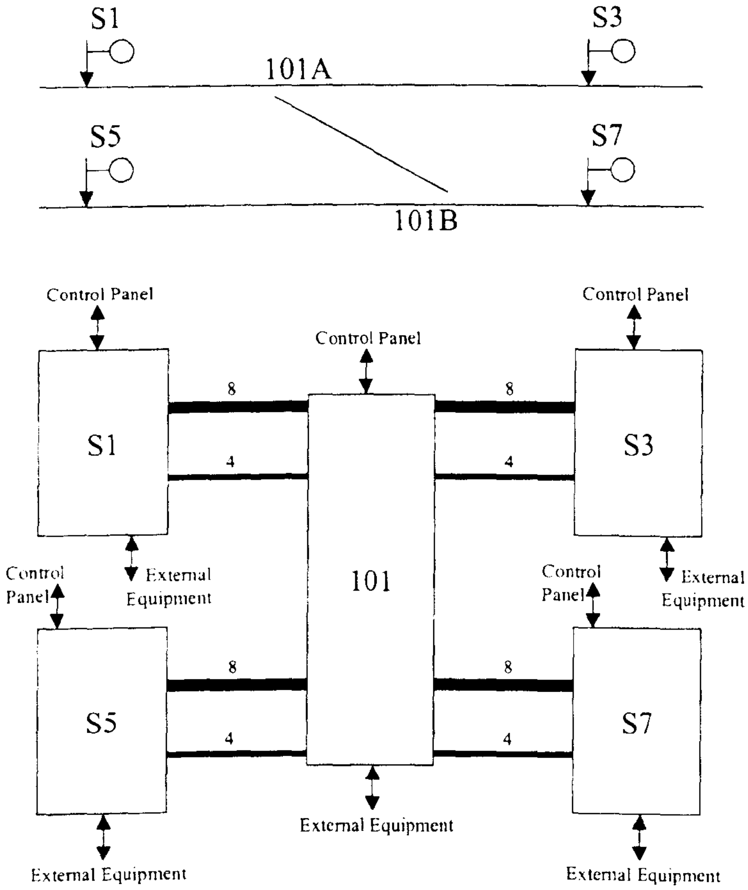
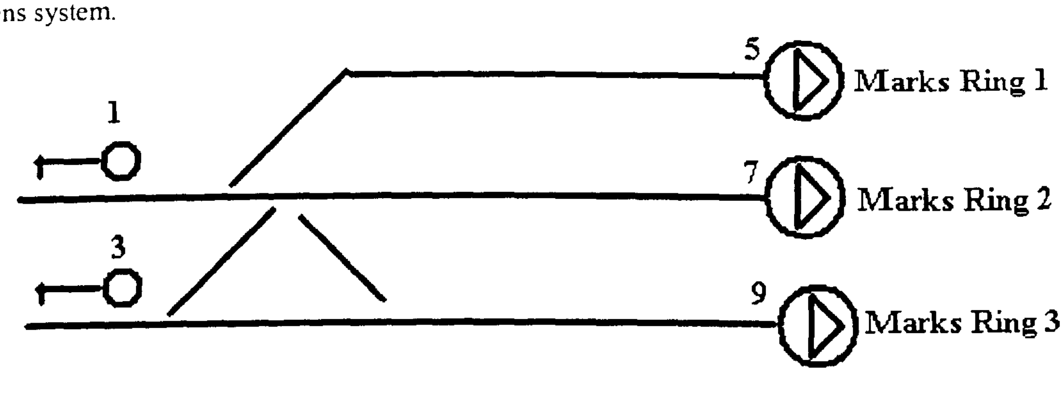


Figure A1: Geographical Interlocking

D. Why they are not used anymore;

SSI worked out cheaper still than geographical interlockings, which lead to the end of their use in the UK. They are still used in a simplified version in South Africa. They have simpler, standard layouts there, so they did not need to have the facilities for looping out. As far as I am aware they still use Spoornet, which is a Siemens system.



Signal 1:
UR for exit 5 connected to Ring 1
UR for exit 7 connected to Ring 2
UR for exit 9 connected to Ring 3

Signal 3:
UR for exit 7 connected to Ring 2
UR for exit 9 connected to Ring 3

Figure A2 Route rings

Kerry Schofield BSc(Eng), CEng, FIMechE (Principal Engineer, Interfleet Technology).

Questions Raised by D. Woodland on 21/01/2002.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I have been involved with railway braking since 1975. I started working on the braking system of the Advanced Passenger Train (ATP). I had a brief spell working on transmission systems, but largely on railway braking. I dealt with a lot of dynamometer testing of brakes (the discs, pads and hardware). I have also had quite a lot to do with adhesion and WSP equipment. I have not had much to do with the control side – although I am fairly aware of it all.

Qu 2. In 'Friction, Wear, Slip and Slide: Braking Systems for Railways' (Fundamentals of Railway Traction Systems, MSc Programme in Rail Systems Engineering, Sheffield University, 2000) you stated that 'Jerk rate is important in braking specifications to prevent people being knocked off their feet and to minimise buffeting. Usually, the maximum is 0.5m/sec^3 for application and stopping at zero' (p9/7). Why is it limited to this value?

I did a bit of digging around after I read that question. My initial reaction has been confirmed by everyone that I have talked to. It is just cribbed from one train specification to another. No one really knows where the values used have come from, but as they have been used before it is generally thought that they must be alright and it is best not to push your luck! Typically values quoted would be 0.5m/sec^3, 0.7m/sec^3 or even 1.0m/sec^3. These figures have to be considered both in the initial brake application and also during blending between friction and dynamic brakes.
One of my colleagues found another reference quoting 12m/sec^3. I think a lot of the difference may be historical. Older measurements would have been recorded on a pen plotter, which effectively had its own high and low pass response filters built into it. These days we can filter the measurements however we want to. It is possible to record instantaneous jerk rates of up to 200m/s^3.
I am afraid that I can not answer your question conclusively. There does not seem to be any real research to support what the limit should be.

Qu 3. You make reference briefly to parking brakes (p9/7), but do not expand on how these work (i.e. are they tread brakes, ...). Unfortunately, I have also found this to be the case with other sources. If you could provide a more detailed explanation, or a reference to a source that does, it would be of great help.

A freight vehicle will have an individual parking brake on every vehicle. A multiple unit will have sufficient parking brakes to hold itself on a 1 in 40, a locomotive for 1 in 30 and a passenger train for 1 in 100. There is a lot of variation in requirements for parking brakes.
Parking brakes would almost exclusively use conventional friction brakes (tread or disk) just applied by a different mechanism. There are three main forms. Spring applied / air released (which is fail safe, in that loss of the train's air supply causes it to apply), hydraulic application and freight wagons that are generally mechanically operated by a separate wheel or lever that mechanically applies a force to the normal brake rigging. Some freight vehicles do have a parking brake that is wholly independent of the service brake.
On the continent it is common to have more parking brakes than we have. That is largely because we tend to operate standard formation trains, whilst they often have carriages left on their own, so each carriage needs a parking brake.
Magnetic track brakes are held off the rail by springs and lowered by air pressure. There is a form of magnetic track brake that uses permanent magnets. At the end of the brake there is also an actuator that rotates a set of alternate North South poles. It then operates like a conventional track brake, but can be used as a parking brake because it requires no power once applied, these are not in use in the UK.
It requires power to apply it, so it is not fail safe.
That is right, although you could make the control mechanism so that it needed air to hold it off. Then it would be fail safe.
The permanent magnet track brakes are manufactured by Oerlikon, now a part of the Knorr Bremse group.

Qu 4. You refer to the fact that track brakes come at the expense of a high cost and extra bogie weight (p9/17). I have seen similar statements in other sources, but I have not found an explanation in any of them. How much more do they cost for both fitment and maintenance? Also, what are the implications of the increased weight?

They weigh about 1 tonne per bogie. A bogie normally weighs 5 to 8 tonnes, so it is a significant weight. The more you increase the train weight, the more energy usage goes up.
Fitting track brakes to a bogie is quite difficult, because they are large.
Track brakes are generally only used down to 50kph due to the sharply rising friction coefficient as speed reduces. If you left them on, you would get very high jerk rates as the train stopped and would also dissipate a lot of heat into the rails.
Track brakes are not used in the UK, except on Tyne and Wear metro as an emergency brake. When Tyne and Wear metro started operating, the drivers thought that the track brakes were marvellous and kept operating them. Their use had to then be restricted.
When I worked for BR Research we did some work looking at the benefits of track brakes for improving adhesion. We found that down to 4 to 7% adhesion, there was a positive effect, but below that there was very little effect on increasing adhesion.

Qu 5. You refer in the paper to 'Voith hydrodynamic brake, increasing drag, permanent magnet brakes and flywheels' as alternative braking methods. I have come across references to each of these before, but never a description of any of them, what can be achieved by them and the advantages / problems associated with them. I would greatly appreciate some explanation or any pointers in the right direction for obtaining this sort of information.

The hydrodynamic brake, or retarder, is a way of putting a dynamic brake onto a diesel unit. They rely on heating of oil. The APT had a hydrokinetic brake. That used a water turbine to generate the braking force. It has almost limitless energy absorption – depending only on how much water you can carry to absorb the heat produced.
We have already discussed permanent magnet brakes.
Flywheel brakes have often been looked at, but never used on UK main line railways. When you want to slow down, you connect your flywheel up to the wheels of the train. The energy is stored in the flywheel until you need to accelerate again. You then connect it to the wheels again and use the energy for acceleration. There were some studies in the 1970's looking at expanding flywheels, with weights on the end of arms that expanded the size of the wheel as it accelerated. This sort of brake is still dependent on the rail/wheel interface and is limited by how big a flywheel you can get on the train – especially since they are spinning at 10's of thousands of rpm, so containment becomes an issue, as does mass. It is a way of moving energy about, but has not proved effective so far.

I have also heard Eddy Current Braking mentioned.

That is an adhesion free brake. The original prototype German ICE train had an Eddy Current brake. It was not adopted for production because of expense and heating the rails up too much.

I have also come across papers about a Japanese Hybrid Rail Brake, that uses both track brake and eddy current brake technology.

Yes, I have heard about that. It is still a contacting brake but with some adhesion free properties.
The Japanese have also been looking at the use of ceramic particles rather than sand for improving adhesion. They can be directed more effectively and therefore use less material than sanders.

Qu 6. How certain are we that the braking rates we currently assume are actually achievable in practice (and under what environmental conditions)?

A new train will have to be brake tested for specified performance under normal dry rail conditions. WSP systems are now also specified, but no targets are set for achieved adhesion.
Traditionally, the signals are spaced with some contingency to allow for poor adhesion, failed brakes, etc. Usually you would have 15 to 20% more at least.

The actual achieved braking rates become of more interesting when you start to think about ATP and moving block.

We usually assume that we can get braking rates of:

Freight 5-6%g for full service and emergency braking

Passenger 7-9%g for full service and emergency braking

Multiple Units 9%g for full service and 12%g for emergency braking

Trams 20%g

How high a braking rate you design your trains to achieve depends in part on the service pattern you will be operating. With frequent stops you need to be able to brake hard to increase capacity. With high speed operation and few stops, that is not such an issue.

The emergency brake is usually identical to a full service brake in terms of brake force. The difference is usually in the control signal. For an air operated train brake, the air pipe is usually held at 5 bar. It drops to 3.5 bar for a full service and is opened to the environment for emergency brake application. The application for full service might take 8 or 9 seconds by this way the emergency application will only take 2.5 to 3 seconds.

On short trains that brake very frequently, some work was done looking at SPADs. It was determined that if a driver mismanaged the brake control on a multiple unit, which stops frequently, there was very little left to stop the train. Enhanced emergency braking was developed as a result, giving nominal rates of 12%g. The extra brake rate is achieved by applying the same brakes at higher pressure. At the moment this is only required for multiple units.

These figures are only nominal. In reality, you are actually getting instantaneous brake rates much higher – perhaps up to 15%g.

Back in the 70s, BR sent a train on a survey of adhesion around Britain for 12 months, covering representative parts of the route in all seasons. The results produced a bell curve with a mean of 0.23 through the whole year. I am told that 1% of the railway would not support an adhesion level of 0.1 (that is the average brake rate achieved after brake build up). If the data only measured the leaf fall season, the figure would be much higher than 1%.

Do you know whether that was in short patches or long lengths?

No I don't. You would have to talk to AEA Technology.

I did some work for Railway Safety on the braking requirements for high speed trains. One of the things that became very apparent out of that was that whereas at the moment up to 200kph we have a braking requirement for the trains, for the vast majority of the time the trains are not utilising all of that braking requirement. It became apparent whilst I was developing this standard that capacity issues make it ideal to operate higher speed trains at maximum braking rates. To run above 200kph, trains would be braking a lot harder than anything does today, and doing it a lot of the time.

As you start to push the braking rate up, you will find that a greater proportion of the railway will not have sufficient adhesion.

Qu 7. What factors limit achievable brake rates and to what extent do these limits apply?

Adhesion, as we have discussed, is the main limiting factor. This can be overcome in part by use of sanders. Initially, one shot sanders were developed to overcome a critical loss of adhesion. After that, auto sanders were developed to allow use more than once. The one shot sander put sand down at 5-6kg per minute. The auto sander puts it down at 1.5 to 2 kg/minute (so it has less effect on track circuits). A lot of testing has been done to optimise how these systems work. AEAT have also developed a smart sander that optimises the sand distribution rate depending on the detected conditions.

You can not rely on sanders, because you can not guarantee they will have any sand left when you need them.

There are other things that can also be done. With a good WSP system, the first axle will slip and in so doing clean the rail a bit. The longer your train, the more significant this cleaning effect will be by the time the rear of the train arrives. Tests have shown that adhesion improves as each axle passes a point, up to about 16 axles. Beyond that number, the achievable adhesion flattens out. Running longer trains would therefore bean advantage, as would braking harder at the rear of the train than the front to utilise this adhesion increase along the train.

As brake rates are increased, you apply larger mechanical forces to train components. You need to design your bogies to cope with that, so that should not be a limiting factor in itself.

Track wear is, to the best of my knowledge, not really an issue with braking unless you have track brakes. With track brakes the track will get damaged. If used as a service brake, they will always go down in the same place and that could be a complication. It shouldn't be a problem if used for emergency braking only.

As you increase the braking rates, you put the wheels into slip more often. Adhesion is the main limit.

There is a firm called Laserthor currently developing a laser system to clean the railhead.

A previous system was developed for the APT, using a plasma torch. It needed a lot of power, was difficult to control and could potentially cut the rail in half if the train stopped! The new laser system does not suffer from these problems. It acts on the contaminant in the rail head and does not effect the rail itself at all.

Qu 8. A paper was presented by Mr E Goddard (Chief Train Systems Engineer of LUL) to the 1997 IEE Electric Traction Systems residential course, in which he claimed that modern microprocessor based train braking systems are designed to SIL 1 at best. He has subsequently confirmed this view to me during an academic interview. I would welcome any comments that you may have as to the justification of this view?

I think that where software is used in train braking systems, the electronics tend to be fail safe and the software is not developed to such a high integrity. I does not really need to be.

We do not allow systems like WSP to act as a whole train system. They act per vehicle or bogie. If one WSP unit fails, the others should continue to operate. This becomes less of an issue as the train gets longer. I think that is the defence for not having systems designed to such a high SIL level.

Qu 9. There are certainly things that you could do to improve the braking performance that might not be very comfortable for passengers, but you could use them only in a real emergency in order to avoid a collision (such as use of high performance track brakes). What are your views on that approach?

In principle, I would rather be bruised than dead, but either way the rail company will get sued!
I think it is a reasonable premise that there must be a brake rate that is uncomfortable but still safe. I expect that the level would be higher than we usually brake, but nowhere near as high as a car with seatbelts could safely brake.

I have heard it suggested that the biggest single improvement that could be made to railway safety would be installation of seatbelts.

Probably, but just try doing it! Other things you could do would be always having rearward facing seats and all seated trains. All of those ideas would be safer and allow higher brake rates, but would they be acceptable to passengers?

Qu 10. There may be things that you could do to increase braking rates that would not be fail safe, but if used in an emergency would actually be more effective at preventing a collision than the current emergency brake systems if they did work. Do you think that such an approach should be considered?

In past experience, if you can come up with a technically possible way of stopping a train faster but can not guarantee it will always be able to do it, it would not be acceptable. There would be a big problem with public perception.

What if you arranged your system so that it was implemented on a per bogie basis (and therefore unlikely to fail simultaneously on more than one bogie), developed a high level of crashworthiness in the vehicle and demonstrated a low probability of a collision if the brake does fail to activate fully (for example by having a braking target of the last reported position of the train in front, knowing that even if it hit a wall as it reported its position it will move some distance further before coming to a stand). Could a probabilistic argument like this be acceptable?

I would be uncomfortable with that, but then I happily drive down the motorway at 1 second separation with less probability of safety. I think that it is worth considering, but I am not sure that it would be accepted. On a closed system, I think there may be more belief in your proposal – a line on LUL rather than WCML. It is certainly worth looking into at an academic level. If the concept could be proved, people may start to be convinced.

Interview by telephone with:

Phil Sharpe (Engineering Manager, Porterbrook Train Care)

Question Raised by D. Woodland on 26/09/2001.

I am interested to know the deceleration and jerk rates used during the operation of the Sheffield Supertram vehicles?

They ensure a full emergency stop from 50mph within 110m or less. That is a deceleration rate of about $2.16m/s^2$

{stopping from 50mph within 110m actually requires a nominal deceleration rate of at least $2.27m/s^2$*}.*

The vehicles each weigh 52 tonnes and use electromagnetic track brakes to achieve the required brake rates.

Tony Sprawl (Training Consultant, Rail Training International)

Questions Raised by D. Woodland on 26/09/2001

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I started as an apprentice diesel / electrical fitter for the BR eastern region at Stratford works. I then became an electrical trouble shooter for BR Eastern region South end Line at Liverpool Street. After that, I went to London Busses as an Auto Electrician for Central Road Services. Then to London Underground

as an electrical fitter on the Northern Line. I was then promoted to technical assistant to the failure and delays inspector for the Victoria Line. Then to Assistant Depot Engineer on the Victoria Line. Then to the training centre as a rolling stock trainer, then divisional rolling stock trainer. I was then attached to London Transport International.
I eventually left LUL and went into teaching for 2 years, specialising in electric traction (city and guilds, B-Tech, electrical, mechanical, pneumatics, hydraulics, electronics). I was then made redundant, so I set up my own business in traction and rolling stock training, eventually joining RTI.
Recent projects I have been involved in include Jubilee Line, Northern Line, Juniper trains, Gatwick Express, Scott Rail, Adtranz class 365, Dublin DART, and South West Trains.

Qu 2. What are the effects of adhesion? What potential is there to limit these effects / by doing what?

Wheel flats are a real problem caused by poor adhesion.
The main parameter that governs the limits of train braking is also the adhesion on the track. The 455s on South West have sanding fitted on the trailer cars for braking. It is only provided for slide, not spin. The Scott rail Juniper trains have the same. Locomotives sometimes have sanding on for spin as well. On multiple unit type trains these days, the axles are all monitored. If one is detected to be slowing up faster than the others a shot of sand is automatically dumped on to the rails to improve adhesion. Normally the brake would also be released whilst that is done. Then when the wheel is detected as turning again, the brake power is reapplied. If it is then detected a sliding again, the slide protection is activated again in the same way. You are allowed 'x' number of activations of this protection during a brake application. After that, the activity has to be cancelled (or there would be no brake force).
On the Juniper, once slide has been detected, the brake force is ramped down to 0kN. The idea is that having dropped the brake effort to 0kN, adhesion should be regained at that point. The wheel's motion is continuously monitored by a microprocessor. The processor then rapidly ramps up the brake application to 60% of the original brake effort and then continues the rest of the application build up very slowly. The idea is to reinstate braking as quickly as possible, but not to maximum force, and then do the final bit slowly to try to avoid the slide occurring again.

Is that done on the emergency brake, or only on the service brake?

A good question, it didn't use to be active on the emergency brake. I think that it is now though. I suppose that it does have the potential to make your brake not fail safe though. I am not sure whether it is active or not.
What they also have now is a 'fifth' axle. It does not really exist. Once the train starts moving, it calculates the average speed of the 4 axles and considers that to be the speed of the 'fifth' axle. When the train goes into braking, the slip protection can not detect the slip by comparing axles if all axles go at the same time. So, it looks at all 4 axles and compares them with the average 'fifth' axle speed just before that moment. That way it can see if they have all started to slide.
Wheel spin is a different thing to slide. As soon as the wheel is detected as starting to spin, they ramp off the power immediately. Then they detect with a microprocessor the moment that the spin stops and ramp the power back up quickly.
The drivers like the braking on modern trains like the Junipers. It is much more responsive.
I read somewhere that they have just developed a new solution that can be sprayed on the track to stop wheel slide. They have tried all sorts. Blowers, Sanders, Solutions, Scrubber blocks. All of them were successful to some degree, but none have solved the problem.

What about track brakes?

If you are talking about trams and light rail vehicles, they use magnetic track brakes. The Euro Night stock (that never actually ran) had magnetic track brakes as well. The Dutch also use them on metro stock.
There is a new type of track brake that has just become available. They all used to be electromagnetic, but one with a permanent magnet has just become available. I am not sure who makes it.

Would that be fail safe then, if it is based on a permanent magnet?

I think that it may well be. It is literally like an anchor. You drop it so that it becomes closer to the track.

Is there any reason why we don't use track brakes here on main line or metro trains?

They are very effective and you stop very quickly. I don't know why we don't use them though
Do you have access to the Railway Gazette, International railway journal? Other countries do a lot of things to control adhesion that we do not, so you may find useful information from them.

Personally, I think that one of the big problems we now have with adhesion is that we have reversed the power to weight ratio over the last 40 years. We used to have a 2000 Horse Power locomotive that weighed 160 tonnes. You now have an 8000 horse power locomotive that weighs 80 tonnes. Immediately, you have no weight to keep control on the track.

Qu 3. Would it be possible to up grade trains with poorer braking systems so that they have performance similar to that of the better systems?

The problem with retro-fitting anything is that it has an enormous cost. The last thing you want to do to any vehicle is retrofit. The 455s on South West are currently being retrofitted with sanding systems. They are 20 years old (so they have about 10 years left). The cost of retrofitting the sanding equipment is astronomical - millions. For a fleet of maybe 80 or 100 trains with 10 years life to go, they have considered the cost worth while for the benefit. So, there is the possibility, but generally it is not worth it.
All train have a specified braking performance. You are not allowed to change their braking performance from that. So you would not be able to change the brake rate without obtaining a new safety approval.
On modern trains, the service brake tends to be dynamic. All of the forces are electrical and only on the motored axles. On the Juniper that means 6 of the 16 axles that would be braked with a friction brake. The performance of the dynamic brake is limited, so if you need more you blend in the friction brake as well. The dynamic brake is not fail safe, so it can not be used for emergency braking. If you isolate a motored bogie, the train's microprocessor detects that and blends in more friction brake to replace the dynamic brake that is not available.
The class 91 has a friction brake on the shaft of the motor. At 140mph braking the axles is not enough on its own, so they have a disc brake on the motor too. As far as I am aware, it is the only one like it.

Did that reduce slip slide problems?

I don't really know. The Eurostar trains have 4 disks on each axle to assist braking. The class 91 with mark 4 coaches had that as well. Twice the usual number of disks.
Basically, there is no reason why you couldn't achieve higher brake rates, but you may have problems with implementing it.

Qu 4. What application delays are applicable to different types of train braking system?

On a modern microprocessor controlled train it is almost instantaneous. With a dynamic brake, it builds up instantaneously at higher speeds, but is not so good at low speeds.
If you use the friction brake instead, the build up will be slower due to delays in the air system, but it is still very quick. These days there is an air supply on every vehicle on every train, so you don't have to wait for air to flow down the train like you used to when you open the brake valve, it is there straight away. These days the air supply reservoir is located right next to the axle. When you operate the valve, the time factor involved is then minimal. If the train was designed badly, with more than 3 or 4 feet between the brake valve and air reservoir, the brake build up delays would start to become much longer. Design is critical to keeping the application delays down.
At higher speeds, you can not beat a dynamic brake. Unfortunately, the power dies out as the speed reduces, so you need to blend in the friction brake to actually stop the train. As it is not fail safe, you can't use it for an emergency brake though.

I have heard that hydraulic braking systems have lower response times than pneumatic ones?

I can give you a pointer on that one. We were involved in the Star project in Kwala Lumpur. The metro system there has a hydraulic actuators in the braking system. That is the first one I have come across. It was made by Knorr Bremse, a German firm. You could try contacting them for information. I think there was an article on hydraulic braking systems in one of the railway journals within the last month or so too, but I can not remember which one.
I am not personally aware of the difference (if any) in application times.

Qu 5. Are there systems that could achieve higher brake rates that are currently used?

I am working for South West Trains at Wimbledon. They have actually reduced the braking performance of their new trains, because it was braking too hard! The drivers thought it was brilliant. They had it for 12 months and liked the way that it stopped. Then they did tests against the specified braking curve and found that it was braking better than specified, so they had it reduced. I could not get to the bottom of why it was a problem that the brake was so good, but there you are.
Have you considered ATO railways? The braking rate with ATO tends to be much higher.

Why is that?

Because the train responds automatically, without the human reaction times. On Docklands Light Railway and the Victoria Line the performance is far better than other lines that use manual driving.

Qu 6. What are the effect of jerk rates on passenger comfort and safety (including maximum jerk rates for passenger comfort / safety?

Somewhere I have the jerk rate figures for the Juniper trains, but I can not remember what they are. I may be able to look them out for you.

The Alstom trains for Singapore Metro have an emergency brake rate of $1.4m/s^2$ and a maximum jerk rate of $0.65m/s^3$. The maximum service braking rate is $1.25m/s^2$

I don't know about limits for comfort or safety.

Qu 7. What is meant by the terms 'service' and 'emergency' braking?

Under UK law, the service brake has to be continuous and automatic (that is the board of trade regulation....). Effectively, continuous means it must be available on all vehicles and automatic operation means that it must come on if the vehicles split.

The emergency brake has to meet the same requirement. It is usually a higher performance brake (1 to $1.2m/s^2$, where the normal service brake is about 0.9). The emergency brake must be operable by passengers & train crew and must be fail safe. That means it must be electrically energised to release. You can not use the dynamic brake for emergency braking, because it is not fail safe.

The service brake needs to be used frequently, with rapid application and release times.

The emergency brake is there for a sudden stop, to be used infrequently and guaranteed to work.

Qu 8. I have heard that the interface between ATP systems and train brakes is usually designed to SIL 1. Is this true?

I can't comment on that. I do not know.

Qu 9. At the moment, we require an emergency brake to be fail safe. There may be things that you could do to increase braking rates that would not be fail safe, but if used in an emergency would actually be more effective at preventing a collision if they did work. Do you think that approach should be considered?

The dynamic brake would give you a better performance but would not be fail safe. In theory, you could do what you are suggesting. Rheostatic brakes are pretty reliable. I see no reason why you could not use them. Regenerative braking would be a different matter – if you were passing a rail gap or neutral section you would lose the supply. It is too unreliable. I would not disagree with your theory, but you would still want something that was highly reliable.

Qu 10. What is the current crashworthiness of rail vehicles?

Modern trains are built like battleships! They have all sorts of protection built in.

Would I be right in thinking that the crashworthiness would tend to be designed to protect the passengers, but not the driver?

No. If you consider the Juniper, the driver's cab is a reinforced cage. It is bolted on to the train. The under frame buffer gear is a big hydraulic ram, like a damper. If you are hit, the buffer goes in like a damper. If that does not absorb all of the energy, the cab mounting bolts sheer off and the under frame moves back underneath the vehicle on slides. All of the semi permanent and auto couplers do the same. They collapse inboard. The idea is that because they are hydraulic cylinders, they absorb energy as they do this. So, all 4 vehicles on the Juniper train absorb energy in this way as they collapse together. The vehicles therefore come together at a controlled rate, with energy being absorbed as they do so. At the end of each vehicle body, there is also now an anti override design. So in theory, the train collapses in a controlled manner, absorbing the energy that would have killed people.

After a collision, there is a lot less damage on a modern train than there would be on an older one. There is a video showing the comparison. Railtrack would have a copy. As the damage caused is all controlled, it can be repaired much more easily afterwards.

You might want to get hold of Rodentum Schafenburg at Hays in Middlesex (a part of the Alstom group) to see what they can provide. The degree of crash protection that you get is down to how much you are prepared to pay for it.

Qu 11. There are certainly things that you could do to improve the braking performance that might not be very comfortable for passengers, but you could use them only in a real emergency in order to avoid a collision. What are your views on that approach?

If you think about an aircraft landing, that is quite uncomfortable. It is better to stop than hit the bloke in front.

You need to be careful about what braking you assume that you can achieve though. If you think about rail oilers, they sometimes go out of spec and cover the rail head. If a train then hits the brake hard as it is passing, the train will not stop. It would go into a slide.

Saying that, I think that you could do it. It would probably be the ethical thing to do if you could save passengers lives. Whether they fall over or not is another thing. You would have to be careful though to make sure that by trying to brake harder you don't just lose adhesion and make the braking worse.

That could be where track brakes would come in.

At the end of the day, we want to stop the train before a collision. You want to throw the anchor out at the end of the day, but you do need to be aware of adhesion limits.

Richard Stanley, MEng, AMIEE (System Modelling Team Leader, Alstom Transport Information Solutions).

Questions Raised by D. Woodland on 11/06/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

I first joined GEC General Signal in 1989 as a sponsored student. They sponsored me whilst I did a Masters in Electronic and Electrical Engineering at the University of Surrey. During that time, the summers (and some extra time) were spent within the company, learning about how it worked and the different organisations within it. The course completed with a research and development project within out R&D department for about 6 months, on a radio location system. Up to then, I had spent some time on basic signalling and SSI systems, in designing a test bed for all of our equipment for training. After the course, I spent a while longer in the R&D department, until it was closed and transferred to Manchester. During that time I did some failure modes and effect analysis on the original ATP system at Marylebone. Following that I did some general work de-bugging the signalling system that was used to control Manchester Metrolink with an ATS and SCAD system, and also the vehicle recognition system (that was used to recognise trams going in and out of the street running section.

Following that, I moved into the IT team, due to the lack of development work going on at Borehamwood at the time. I spent about 3 or 4 years in that team. Then in 1996 I was asked to move back into Engineering on the TCS project, as the TCS performance engineer – looking at RAMS performance and lifecycle costs. As the project has grown, a team has built up around me to do that work. So, we have been involved from the very start looking at the performance, how the system works, some of the operating principles and the concept of the system design.

Qu 2. The terms 'Signalling', 'train control' and 'railway control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system?

Signalling I see as the low level protection, giving information on the state of the railway ahead to the driver so that he can take the decision as to how fast and how far he can go.

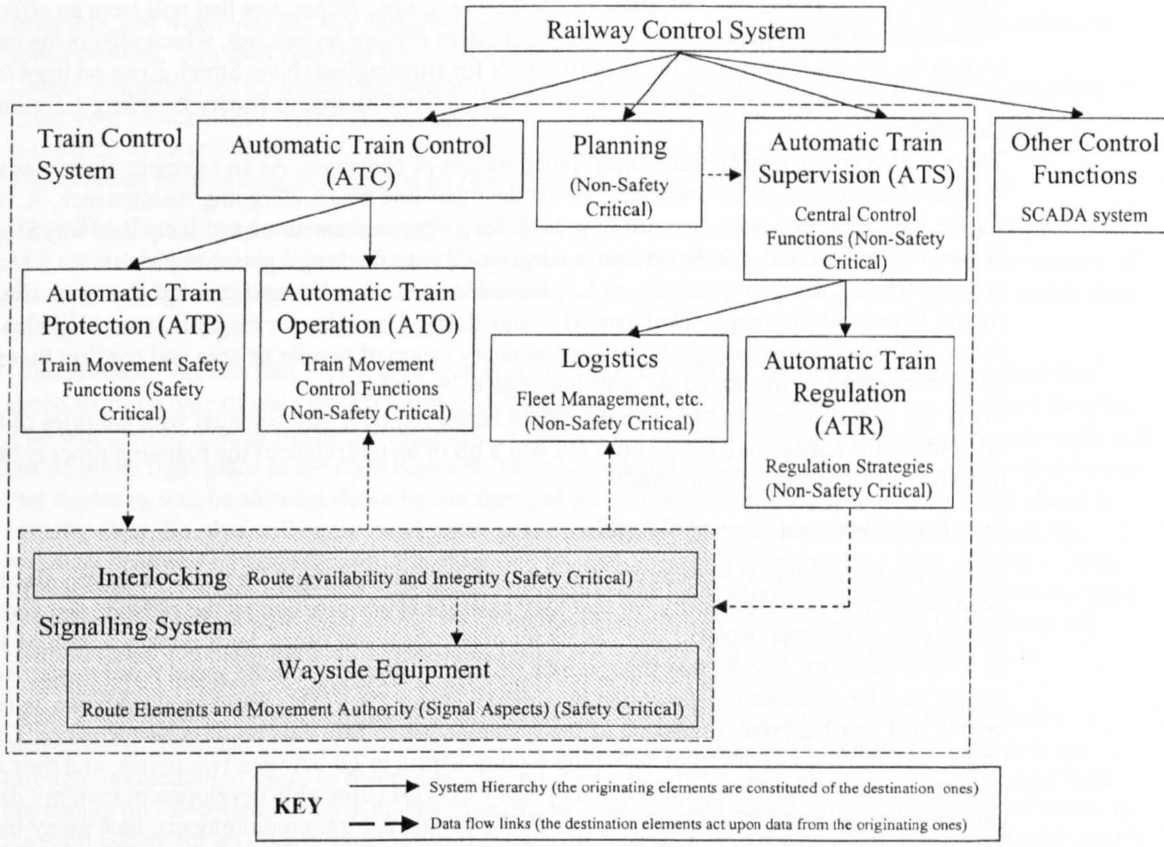
Train control implies some aspects of centralisation and more control of the train (such as controlling it to remain within acceptable speed limits). That would include ATP functions, whereas signalling need not necessarily do that.

Railway control I see as much more than that. The functions of ATS or network management. Not necessarily protecting the train, but directing where they go and regulating their running against the timetable. Providing information on the actual train condition, system failure recording, all of the things that you need to manage the railway.

Qu.3 Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

Regulation I would take as one aspect of supervision, making sure that trains run in the right order to a pre-defined timetable. Providing information to the train to let it coast or tell it to speed up, in order to keep it to the timetable. Also decisions on the order that trains are routed in, to support the overall timetable working.

Automatic train supervision I see as moving more towards network management. There are other functions in ATS besides ATR. Questions of fleet management. Cancelling or turning trains back. General SCADA systems would be outside of train control and a part of railway control.



It might be interesting to consider dividing the system up into supervision, control and protection of signalling. What happened in the past was that the supervision was the setting of a route on the signallers panel. Now though, you could see them all as separate functions or systems.

Qu 4. How would you define the capacity of a Railway system (and what would you consider to be appropriate units for your defined measure of capacity)?

The simplistic way of defining it is trains per hour, or headway. However, that assumes that all of your trains are of the same type and there is a question as to whether you have optimised the railway for a certain train type or not. As soon as you impose fixed blocks on a system, it can only be optimum for one type of traffic. So if you have varied traffic (like on the West Coast), you can not express what capacity is in purely trains per hour. Your capacity is then entirely a function of your timetable and the mix of traffic that you want to run. The actual signalling and ATS is of secondary importance to that. It is a question of how many trains you have got, what there characteristics are, what you want to do with them and how much railway you have to run them on.

One of the useful ways we have looked at this problem is to take a standard train type with a standard stopping pattern for a line and to then express everything in those terms. If you then run a different type of train, you can show how much more than one train path it takes. If you look at a train graph, you will see lots of void space around a slower train. That is wasted capacity. If you are charging by train path capacity, say, you can then actually charge for the extra train paths wasted.

When you do an analysis of the timetable, you can tell how much traffic goes between two points. The mix of traffic affects it so much, that you really need to account for that as well. If you say that you have a theoretical 100 train paths a day and you are using 91, you could say that you are 91% loaded. What you don't know in that, however, is the stopping patterns. You may really be 120% loaded and causing inherent delay to the timetable.

One of the problems we have at the moment is that there are no margins in any of the definitions that we are using.

One of the most important things is the stopping pattern. That really brakes everything. If you stop one train and the next one goes straight through, you have immediately wasted a lot of capacity.

There is then the point that flighting trains of the same type and stopping pattern together in order to maximise capacity. However, that might not meet your customers demands.

So, trains per hour between two points is quite a useful measure, but you must define that standard train.

There is another factor, which is the difference between technical and operational headways. Technical ones you can calculate. The operational ones than need some margin. The tighter your technical headways get, the higher your % of the operational headway needs to be margin. There needs to be a measure of resilience and what is acceptable.

You really need to look at the complete timetable planning process, to get an idea of how they go about it.

There are things like driver variation and defensive driving techniques that will have an effect. Different Train Operating Companies have different defensive driving techniques, which effects the capacity. Defensive driving was one of the biggest issues for Birmingham New Street. Even on lines where it does not have such a significant effect, defensive driving from Euston to Rugby is adding 1.5 minutes or 2% to the run time.

There is also an interesting issue in degraded modes of operation. As an example, there was recently a section of line on single line working whilst the other line was undergoing maintenance. A 16 vehicle freight train arrived, under instruction to look for a signalman with a hand lamp half way along the line. When the train arrived, the driver saw a fairground, with flashing lights everywhere. As a result, he was not sure where the signalman was, so he proceeded at caution (10mph) through the work site, when he should have gone at 20mph. That caused a huge delay. The rules for single line working also say that the driver of the first train through has to stop at every group of people he sees and confirm that they are aware the section is in single line working. Again, big delays.

There are a lot of issues in how you develop a timetable, how you come up with the rules of the plan and do your performance analysis. I could tell you a bit of an overview of the planning process behind timetable development if you want?

Yes please, that would be very helpful

If you take the railway system as a nodal network with timing point locations at certain places, which can be mandatory or non-mandatory, so that they exist for every junction on the railway. Between pairs of timing point locations (tiplocs) you have a set of sectional run times. So if it takes a train of a given type so long to run from A to B, that time will be different depending on the speed it has to start and finish the section at. The runtimes are generated by a model called TRATIM, which is also known as the Derby model. It is a rather poor simulation of the system using certain data on train performance. Using these times, they build up a pattern of how long it takes a train to go between two points, and they use that to plan the journey time for a train. Then they use PROTIM (timetable development system / database) to generate the timetable. They have a list of all the contracted train requirements, how many trains per hour. They have to feed all of those in. They usually start at the Coventry corridor, Birmingham New Street and Leeds. These are the pinchpoints on the network. They start with them and work out. They are 100% loaded, so if you don't start with them, you have no hope of resolving any conflicts that arise when you do get to them. They tend to start with a standard hour pattern, then build that up to 2 hours, 4 hours, getting bigger until the timetable is complete. They then simulate that run with the VISION simulator, which highlights any routing conflicts that they might have.

The problem is that the train routing is effectively done by hand. The train planners who initially generate the timetable have a view of how the trains are going to run. That is then lost during the transfer to VISION. The people running VISION then have to make their own assumptions about how the trains are running. Then if you go to visit an actual signal box out on the line, you will find something different actually going on on the day. There is some research into what actually goes on at the moment, which will be giving an algorithm to the NMC project.

Theoretically, you should then have a timetable that meets your capacity and journey time requirements, and has been simulated to show that it works. Then they go to the timetable conference with the TOCs, which is when it all gets agreed and signed off by the TOCs and becomes a timetable.

During the simulations, they have an infrastructure model of the railway within a tool called TRAIL (a monti-carlo simulation model), which is a statistical failure model of the railway run by Jardine associates (who are a company based in the oil and gas industries).

They can associate an impact with each failure. They take the timetable and do a monti-carlo simulation of failures over 30 years, overlaid with the developed timetable, and work out the expected minutes delay and the apportionment of whose fault the delays are. It can also work out how much the penalty clauses in contracts would make the delays add up to in pounds. From that you can get an estimate of the punctuality requirements, and whether or not they will be met by the timetable.

At the moment on west coast, the sectional run times used in the timetable development are tight to the technical headway. On the southern zone they add a 5% allowance in. Dutch railways add 7%. For some reason, west coast does not add anything. That means that the timetable is being developed to times that will be almost impossible to meet.

Qu 5. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?

Performance impact and cost.

Looking at the different types of ATP that we have, the difference is really in the methods of communication to the train, from a performance viewpoint. When you can communicate with the train. With a radio system you effectively have continuous infill, so the train can react as soon as a piece of track becomes free. With discontinuous systems, be they loops or spot transmission, you have major performance impacts in operation. Everything can be designed to be fine when things are running

smoothly, but when you get any perturbations in the system the trains no longer run to their optimised speed profiles. You then find that the service gets a lot worse very quickly.
Unless you mitigated by improving performance in other parts of the railway, ATP systems are going to make your performance worse. Rather than being seen as an advantage, they are therefore seen as something to be criticised and blamed whenever anything goes wrong.
The question of installing anything on an operational railway is going to be a problem as well.

When Tim Cusk did a presentation to the IRSE Younger Members' Section, he said that the capacity of TCS level 1 would be much worse than that of conventional signalling, due to the intermediate nature of the communication system. However, he also said that the capacity of level 2 would be much worse than level 1.

I can probably explain that. For level 1 it is fairly obvious that if you are running at the designed line speed with the type of train the system was designed for, the balise placement will be optimised for you. You will then be able to run as conventional. However, the moment you get a perturbation, your train will not be in the right place at the right moment. The discontinuous nature of the communication then means that the trains will be slowing down before they get the message from the balise that the track ahead is actually clear. So, that will cause much more perturbation. We have measured that, and it increases technical headway by something in the order of 11% and operational headway by 15% in level 1. Also, without infill in certain cases, for high speed (125mph) running in level 1, the trains would actually need to start braking before they reached the YY. You therefore effectively reduce the line speed from the basic 125mph. On plain track with minimum signal spacing, the trains would only be able to do something like 102mph. So you have to have infill for that arrangement.
Level 2 will have an impact because it is an overlay on conventional signalling. The RBC gets the instruction of a change in signal aspect at the same time as the signal itself. The RBC then has to go through its calculation cycle to determine what the change in movement authority actually is, and then depending on the communication update system you have chosen (polled communication to all trains, or the train asking for an update when it wants one), there will then be a delay before it is transmitted over a radio communication network. That may have a nominal value of 7 seconds, but if you have lost communication and have to reconnect, it could be 30 seconds or more.
At the moment, we will have a permanent connection (unless it is lost). The system will operate like a mobile phone network. Later on, we will use packet mode GPS, you will still have a delay on distribution then, but the distribution tail will not be so long. You won't have the tail of reconnecting. We have calculated a 15% increase in technical headway for level 2, worse than level 1. But the continuous nature of the infill actually improves the resilience. If you take a 7 second guaranteed sighting distance for conventional signalling, the continuous communication in level 2 would actually improve the operational headway over level 1.
If you want any detailed parameter based performance measurements, they are all in our tools that David Fisher has created. I am sure he would be happy to take you through them.
If we used an on-request update system, rather than a polled one, then we would have to assume that everything could happen at once. There is the possibility of all trains wanting an update at the same time, which means you could get bursts of activity (particularly after an incident) which would load the system and slow it down. That would change the impact of level 2 from 15% on headway to 22%.

Qu 6. I would welcome any views that you may have as to ways in which Automatic Train Control systems could best be optimised to create the greatest benefit for UK railways in both the short and long term?

Looking at the long term, the system should be designed for life cycle costs. Not just the costs of designing and installing it, but also the penalty charges for delays consequential to having the ATC there. AT the moment, we tend to only look at the short term costs, which will impact the railway for the next 30 years. Do we really want to do that? We are not always being sensible in the way we budget for systems.
There is also the question of looking at fixed and moving block. Our early analysis on the TCS project showed that the improvement you get from having moving block, which effectively optimises the railway for any train pair allowing all trains to run at minimum headways, overcomes the limitations of fixed blocks. You can get better capacity and higher line speeds with a fixed block system by shortening the blocks, but that gets expensive and introduces unreliability. Moving block would therefore be a big benefit.
Looking at improving the railway as a whole, the removal of lineside equipment (anything in a red zone that may need to be maintained) and building resilience into the system (so that single failures do not stop trains) would be of great benefit.
So, it all comes down to life cycle costs and protecting the revenue stream for the railway.
How reliable the system looks to the customers, whether they can rely on it getting them from A to B on time, is also important.

How much did we do on our analysis of moving block to consider the effect of complex junction arrangements, as compared with plain line running?

I don't remember off hand. We definitely did some, because we had the concept of a junction block. It may have been in the work that we did for black diamond. I suggest you speak to Dave Fisher.

Qu 7. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

Firstly to get it approved! Then making it easier to install on both trains and track. Why can't we have a hand held unit on the train, that would make it easier to install.
The difficulty of installing equipment onto a train is a big issue at the moment. It is very invasive and requires a train to be taken out of service, which the TOCs can not afford. You can also not get much access to most operational railways to install equipment on the track. Things which can be installed away from the lineside are good.

Qu 8. The introduction of higher speed operation, higher capacities or changes to track layouts may all offer operational benefits, will influence the commercial viability of projects and will also affect safety. How do you think the trade off between safety and operational flexibility can best be managed?

The introduction of the higher speeds and capacities means that you get a reduction of inherent safety in the system (that you maybe had through timetabling). That does need to be measured somehow.
You might think that the timetable is designed to allow all movements on time without conflict (which would be safer). However, we have been running some analysis of current timetables in order to check the tools that we have developed, and have found that the timetablers actually plan trains to conflict at junctions on the assumption that late running or signaller action will actually prevent them arriving at the same time. So the railway is not actually planned to be safe. They are relying on the signalling to protect the trains from each other and prevent accidents. That means that risks are being planned in to the timetable. We found that at Nuneaton in the current timetable there are 53 possible ways to move through the tiploc, and of those there are 805 possible interactions of possible conflicting moves through the station. The 2000 winter timetable broke the rules of the plan were broken 816 times. That is 1 in 3 trains that were planned to conflict with each other. I wonder whether that is factored in to our current view of the safety of the railway?
We currently actually have unplanned safety because we do not saturate the railway. However, we are moving towards that saturation soon. This is going to impact the safety. The probability of collisions after a SPAD will increase. We therefore need to consider what we are doing to protect these trains. That means that fitment scope is important in the trade off between safety and operational flexibility.
Some of the things that were planned on west coast, such as the removal of grade crossings, have now been removed from the plan as a part of cost reductions. So we do need to develop an overall view of the safety of the railway, show what the new capacities are going to mean and therefore get a better idea of what we need to do to optimise the balance of safety and operational flexibility.

Qu 9. What information do you think:

F. Is required by a driver in order to perform the tasks asked of him

The driver needs to know whether it is safe to proceed and the condition of the line ahead (whether through cab displays or lineside signs). He also needs to be able to see what is happening outside of the cab, observing problems, vandalism, trains in other directions. That observation is important.

G. Would be beneficial to a driver to assist him in the performance of his tasks

It would help the driver to drive if he knew how close to timetable he was. Many of them have lists of when they should pass a particular point and have knowledge of how long it should take. Real time information on how well they are doing against the timetable, and whether they are being regulated differently would help the railway run better.

C. Would be detrimental to a drivers ability to perform the tasks required of him

Any trivial information that the driver does not actually need to know, and may therefore distract him from his driving. That would be a problem.

Qu 10. How do you think that your answers to question 9 would change if the system included ATO?

Baring in mind that even a 1 minute delay in a critical area (such as a junction) on the 2005 railway will be catastrophic, with penalty charges for delay and consequential delay at about £180 per minute you could easily lose half a million pounds. You need to provide the driver with information to minimise delays if the ATO does go wrong.
You also need to make sure that the drivers task is still interesting, so that he does not neglect looking out of the window.
ATO would have a performance benefit of removing a lot of variability in driver performance, which would in turn improve capacity.

Dave Stevens (Traction and Rolling Stock Associate, Rail Training International), Barry Garner (Traction and Rolling Stock Training Consultant, Rail Training International) and Ben Fry BSc, MBiol, CBiol (Marketing Manager, Rail Training International).

Questions Raised by D. Woodland on 26/09/2001

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?

Dave Stevens: My background has been mainly in training. I did work on the railway for a short time as a commissioning engineer for new suburban rolling stock , but then in 1977 I became a college lecturer at a further education college. There were a lot of railway students there, so most of the timetable was built up around teaching them. I left there in 1996 and have been self employed since then. Most of the time I have been working on obtaining European funding on ADAPT projects. Then Last year I can to RTI. Over the last year I have been working on modern rolling stock, putting together training manuals for the new automatic train being built by Alstom in France for the Singapore North East Line.

Barry Garner: I was an apprentice on the railway, working at Wolverton and Illford. Then I went to work for Newham College. I have also worked for First Great Eastern and have now been with RTI for 3 weeks.

Qu 2. What is meant by the terms 'service' and 'emergency' braking?

- **What are the perceived purposes and requirements for each type of braking,**
- **What is the difference between the current performance expected for 'service' or 'emergency' braking (both achievable rates and application delays),**
- **What is the justification for this approach and,**
- **What is the potential for benefit to be gained from reconsidering the current approach?**

Ben Fry: The legal requirement in the UK is for the emergency brake to be 'automatic, continuous and fail safe'.

Dave Stevens: There is a bit of a hang up about fail safe braking. After all, historically if you lost the braking signal (whatever that might be) then you potentially lost all of the brakes on your train. On a modern train though, everything is very modularised. You are unlikely to ever lose all of the brakes. It could be designed down to a per axle basis.

Ben Fry: That is how it is on the Rotterdam Metro. It has all wheel traction and regenerative / emergency braking. The emergency brake uses that electric brake.

Dave Stevens: If all axles were braking through their own rehostatic system, none of them would be fail safe, but what level of fail safe are you looking for? The likelihood of all of them failing simultaneously would be minuscule. I think that you would always need to have an element of fail safe in your emergency braking system, but to what level?

Ben Fry: It would be very difficult to get a vehicle approved or to justify the operation of the brake in an enquiry, if you compromised on the fail safe principle. After all, it is a legal requirement at the moment.

Qu 3. Are there braking systems that could achieve higher brake rates than are currently used?

Dave Stevens: That has been a theme for a long time. There is even the potential to get completely away from the friction brake by pumping hydraulic fluid around a controlled system and converting the trains energy into heat.

Barry Garner: That would be similar to rheostatic braking in concept

That would not be fail safe though

Dave Stevens: No it would not be fail safe and it would also be unable to stop the train because you would run out of pumping action before you actually came to a stop. You would need a friction brake on top of it

Ben Fry: That is always a problem with dynamic brakes, whether electric or hydraulic. They are very good at high speeds, but not at low speed. Friction brakes on the other habd tend to be the reverse of that, more effective at low speed. Then there is the block brake that is more effective at very slow speeds.

Dave Stevens: If you are talking about very high speeds then the delays that occur for the valve to open to allow air into the brakes and for the air to build up to apply pressure, must mean you travel a long way before you get brake application. You could really do with a quick one step application with low application time.

Ben Fry: I would question that. The track brake tends to be used on light rail vehicles such as trams. It is used as an emergency stopping device at low speed. It has a dumb control – on or off. I think there would

actually be benefit in looking at a more actively controlled track brake that could be applied gradually to increase the drag of the train and slow it down in a controlled manner.
Ben Fry: A lot of effort has been put into developing technology to brake the wheels, but you get to a point where nothing else that you do to the wheels will get you any further with your braking because you are limited by the adhesion between the wheel and the rail. So nothing else you could do to slow the wheels would help. You actually need to look at other ways of slowing the train.
Barry Garner: On racing cars, they increase the friction between the vehicle and the road by using down thrust to increase the pressure. That is something that Railways have never done.
Ben Fry: That is the difference between a normal train brake and a track brake. It is an attempt to glue the train onto the track.

Qu 4. What application delays are applicable to different types of train braking system?
Ben Fry: It takes a finite time for a brake application to build up to full strength. I think that is usually measured in a small number of seconds. If you look at the design of conventional pneumatic disk brakes, there are so many components in there that I am sure there is no way to significantly reduce the build up time. A lot of effort has already been put in to doing that.
Barry Garner: If you used hydraulic braking systems instead of pneumatic systems, would that make the application any quicker?
Ben Fry: I don't think so. Lorries use pneumatic brakes and have far quicker brake application than cars that use hydraulic brakes.
Barry Garner: The Advance Passenger Train had a hydrostatic brake. A water pump brake. I don't know how successful it was. It had a friction brake as well, but I think that the water pump was provided to have a quicker application rate.
Ben Fry: In terms of propagation through the train, an hydraulic system would be quicker because it would not compress like air does. I think that one of the problems of build up in brake application on a train is because you are using compressed air.
Dave Stevens: They have got rid of a lot of the delays due to that cause on modern trains though. The whole of the brake initiation is now an electric signal that may feed half a coach worth of equipment that has its own air compressor and pneumatic system locally. When the driver moves the brake lever, there will be a delay in his act of doing that. Then there will be a delay for the proportional electrical signal (but that will effectively be instantaneous right up to the bogie). The proportional electrical signal then has to be converted back into a proportional air signal. That proportional air signal then goes into a valve that goes into an air valve to let the main air from the brake cylinder go through to the brakes cylinders. The delays that we have today all occur in that local air control. In order to minimise that, they tend to fit the brake actuator for the air control right near the bogie, so that the air propagation delays are minimised.
Barry Garner: you could reduce delays by finding a way to control the main air valve directly from the electronic signal, rather than having an intermediate air activated stage. To go for a digital valve.
Dave Stevens: One of the reasons for having an intermediate air signal is because it has to account for passenger loading. The passenger loading signal comes from the air spring and is already in air form. It is therefore easier to combine the drivers brake demand and passenger loading control as an air signal.
Ben Fry: The passenger loading compensation signal is real time.
Dave Stevens: Yes. If you could turn it into a digital signal, you could have it input to a digital air valve along with the brake demand signal. The valve to the brake would then open proportionately to the two signals. That would possibly reduce delays.
Barry Garner: Shorter air pipes and more compressors would help to reduce delays too.
Ben Fry: Design has gone the other way though in recent years Hasn't it.
Barry Garner: It has, yes.
Ben Fry: Design has been reducing the number of compressors, presumably for maintenance and weight reasons. In some cases there is only one compressor per train.
Barry Garner: I suppose that you could speed up the flow of air by increasing the air pressure. Why do we only work at 10 bar?
Ben Fry: Presumably, because as you increase the pressure, so you increase the significance of leaks.
Barry Garner: Then you would need more compressors.
Ben Fry: What we are talking about here is the distributed traction and braking, but distributed traction and braking seems to be going hand in hand with the reverse when it comes to supply of the compressed air. It is now coming from a smaller number of compressors located further away from the brakes, whilst everything else is tending to be done per bogie now.
Dave Stevens: Very localised now.
Ben Fry: Yes, very localised. All that really goes the length of the train is signals on the data bus that address a device per axle.
Dave Stevens: The first service brake these days tends to be an electrical brake rather than the air brake. The one I have been looking at for Singapore doesn't have all axles motored (which would be the ideal situation). What it does is assess the demand for brake and set up the electric brake on the motor coaches

so that the trailer vehicles don't have to have any pneumatic brake on them at all. In terms of speed of application, most service brake applications do not have to work on air application and are therefore much faster.

Qu 5. What are the main limitations on being able to increase brake rates?
Dave Stevens: As far as I know it is adhesion.
Barry Garner: It is metal on metal, if the brake rate is too high, you skid.

On the main line railway you typically achieve around 0.8m/s² braking, but on LUL they achieve 1.0 to 1.2m/s².

Dave Stevens: The train for Singapore MRT is 1.25m/s².
Ben Fry: There is, or at least was, a department of transport specification for brake rate at 0.9m/s² Most trains now can achieve 1.0m/s²0 and with the emergency brake is usually rated at about 1.2m/s².

Qu 6. There are certainly things that you could do to improve the braking performance that might not be very comfortable for passengers and may even cause damage to the track, but you could use them only in a real emergency in order to avoid a collision. What are your views on that approach?
Ben Fry: If you used a rail brake, like the new units in Rotterdam, the brake rate is 3.2m/ s². That is only designed for use at very slow speeds though. The drivers on the metro are all taught to use the rail brake as an extra emergency brake, if the emergency brake alone is not going to be able to stop him in time. Then he presses the rail brake panic button. He is not allowed to use it normally and its use is logged. You could also consider that with braking the wheel, as soon as you lose adhesion and begin to slide, you start to seriously damage the wheel and railhead.
Dave Stevens: If you applied more effort to slowing the train rather than the wheels, you might reduce that as well.
Barry Garner: If you compare the shape of the wheel to the rail, how much of the wheel is actually in contact with the rail? Perhaps you could improve adhesion by increasing the contact area?
Ben Fry: The profile of the wheel has been developed especially to improve stability and improve efficiency. If you think about that logically, you don't really want to change that to help with braking because it would undo all of the effort that you have already put in. You need to find a different way of slowing the train.

Qu 7. What potential is there to limit the effects of adhesion/ by doing what?
Barry Garner: You can use sanders. First Scot rail had them, then southern region and now great eastern have them too. It increases grip when there are problems with leaf mould.
Ben Fry: Scott rail's Alstom Juniper trains have sanders on the leading axles. They are automatically triggered by the wheel slide detector. There is also a sanding button in the driving cab. They are not intended for improving traction / preventing slip.
Dave Stevens: There did not use to be a problem in the days of cast iron tread brakes. They cleaned the wheel, which helped reduce adhesion problems.
Barry Garner: You got less flats with the old block brakes too, as a result of that.
Ben Fry: They have also reduced vehicle weights in recent years, which also reduces your adhesion.

Does the use of sand cause any additional rail / wheel wear?

Barry Garner: No, I don't think so. It reduces wear by reducing slide.

Tam Taskin BSc (Hon), MSc, PhD (Project Manager – ICONIS UK Customisation, Alstom Transport Information Solutions)

Questions Raised by D. Woodland on 01/05/2001.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career within the railway industry?
I spent a long time at University. My first degree is in Electronics and Telecommunicarions. I did a Masters in decentralised control systems and then a PhD in simulation and modelling on railways. Then I started with GEC Alsthom. Then LUL, then Alcatel and now back to Alstom. I am also doing work for Thameslink 2000 Programme and Infraco BCV. I see my role as a "middle man" who understands the railways requirements and operational needs and turn them into technical solutions. I am currently concentrating on control systems and ATS/ATR.

Qu 2. The terms 'Signalling' and 'train control' are often used interchangeably. How would you define each of these terms, and what difference, if any, do you believe there is between a 'signalling' system, a 'train control' system and a 'railway control' system?

Train control covers signalling but has a bigger context than signalling. I find it 'unusual' when people refer to a conventional fixed block system as a train control system or to Alcatel's radio based moving block system as a signalling system. I would call it a Train Control System if an ATP and / or ATO is involved as an integral part of the system. So, the Central line for example would be a 'train control system with fixed block signalling'. The term signalling include interlocking and any trackside – sometimes trainborne - indications associated with it. The term Train Control System normally suggest a more comprehensive solution which also involves a trainborne ATP (and ATO) with associated trackside components.

Railway control system makes me think of the ICONIS type system, with overall control on the railway where this control is exercised through an underlying signalling I would assume the term to therefore include ATS and Traffic Management.

Qu.3 Two other terms that are often used interchangeably are 'Automatic Train Supervision' and 'Automatic Train Regulation'. How would you define each of these types of system and the difference between them?

Railtrack used to refer to Train Regulation as centralised control of the signalling system both by automatic and manual means. This term is now less frequently used – it is replaced with Signalling Control with/without Automatic Route Setting.. The European terms CTC - centralised traffic control – or Traffic Management System – mainly in the context of ETCS/ERTMS – also gained acceptance

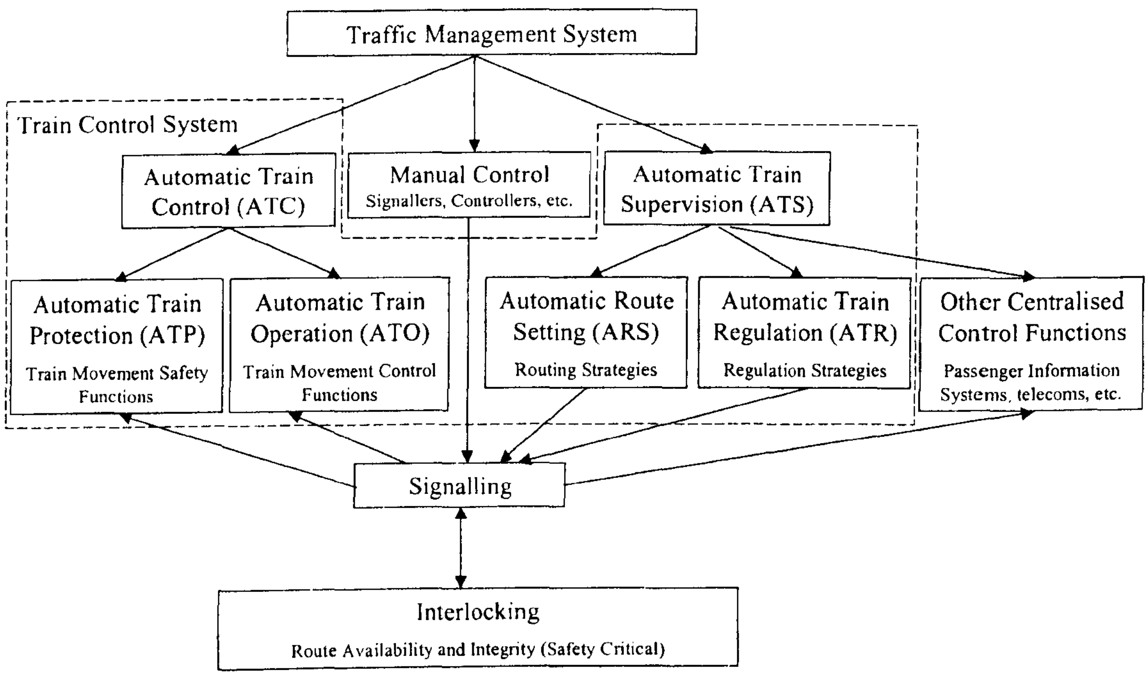
LUL have always used automatic train regulation to refer to the automatic control of separation of trains to provide a regular service.. In this context, ARS is a part of ATS solution.

Automatic Train Supervision used to define the centralised control of train service. ATS systems usually integrate additional functions such as passenger information, telecoms network, etc. It is a generic term for the control of the whole system .

ATS would include ATR, but not ATO or ATP. It was first used in the context of an automatic railway to be the components of the control that were not ATO or ATP. So ATS is the centralised control function – the control room systems and distributed components: Use ATS instead of for example CTC indicates the presence an ATO and ATP in the system.

ATO + ATP = ATC. ATC+ATS+interlocking = total control of train movements on a railway.

As mainline systems do not have an ATO system, mainline operators do not normally use the term ATS. Railtrack use a VDU based 'signalling control system'. They also use NMC interchangeably to describe either the control centre or the control system. Under NMC is IECC type systems which are referred to as signalling management centres. They use these terms instead of ATS. I prefer the term Traffic Management System.



Qu 4. How would you define the capacity of a Railway system?
From a metro viewpoint, it is passenger throughput that counts. That is what you should optimise. It will determine how long your trains need to be and how frequently and regularly you need to run your service hence what your operation headway and service pattern should be.

Capacity is a combination of service pattern and what the signalling system can support in terms of train throughput.

You are also trying to achieve minimum journey time for your customers. So the high capacity system is optimised when most passengers are having a minimum journey time through your system. Increasing tph but travelling at lower speeds may not do that.

On a metro system, regularity is a good indication / measure of maximised capacity. A system that bunches trains together and leaves big service gaps may be able to achieve a high train throughput and carry large numbers of passengers, but is not achieving the maximum capacity.

Therefore, what determines capacity not only the signalling systems designed capacity, or headway, but also how you control the regularity of your train service and plan the train destinations in accordance with the passengers requirements.

On main line applications, you would probably focus mainly on train throughput and regularity of that throughput that gives capacity which in turn relates to passenger throughput. The granuality is higher than a metro system, service frequencies are lower and other concerns in addition to passengers has the taken into account (freight, non-homogeneous fleet, etc.) but the main definition holds. A high capacity system is the one that carries maximum number of passengers from A to B in the minimum amount of thime.

On the Central Line service regulation, we concluded that we needed to optimise for passenger-seconds. Looking to minimise passenger journey times whenever a perturbation occurred to the service. {see report provided}

Qu 5. What would you consider to be appropriate units for your defined measure of capacity?
Have you looked into LUL's key performance indicators? Some of them may be useful in terms of measuring service quality. I equate capacity to the quality of service. There is a perceived capacity as seen by the passengers. That is how a metro should be judged. How long they have to wait, crowding levels, etc. You should consider how an ATP system can help to run a more regular service for the passengers. Moving block systems are a good example, if you don't control it with ATS, you can have bunching of trains, resulting in a less regular service and lower passenger throughput, despite you theoretically higher capacity. How you control junctions will also have an effect on this. It is not just your signalling system, but how you control it that determines capacity.

Higher throughput of passengers or trains with a more even distribution gives you higher capacity, so the measure needs to consider both factors.

Qu 6. From your perspective, what would you consider to be the main issues to consider when introducing an ATP system onto an operational railway?
Testing and comissioning strategy. Keeping the possessions to a minimum. A pilot installation on a remote section of the railway or on a test track that represents the operational environment sufficiently is also pre-request as this will lead to better understanding and resolution of signalling principals and operational rules issues, safety case and product approval issues, technology issues etc. A pilot installation is particularly important for introducing ATP to Railtrack network where ATP will be integrated with and interoperate within the existing infrastructure.

Try to reduce the use of common equipment between the conventional signalling and the ATP. So if they both use track circuits, it becomes difficult, but if one of them is transmission based, using leaky feeders, to determine the position of the trains, it enables over and back testing. That makes the introduction and testing much easier.

On the DLR for example, it was easier to introduce the moving block system because it used different components to its predecessor.

Qu 7. As an additional point to this, I would welcome any views that you may have as to ways in which ATP could best be optimised to create the greatest benefit for UK railways in both the short and long term?
In an LUL type environment, where the trains are uniform, it makes sense not to run trains with only some of the fleet fitted. You are therefore looking for a uniform change.

On Railtrack, it is a different thing all together. You have to run mixed fleet operation.

TPWS is a risk reduction system. It does not eliminate the risk. It may be cheap, but it is not necessarily the most cost effective way. I think that the most cost effective way, on a scheme like Thameslink 200 that needs ATP with low headway, it makes since to have a system that will monitor trains on the approach to stations to reduce variation in driving technique and the need for defensive driving. That would help to make braking curves more predictable (under an ATP profile), and thus aid capacity. So station areas are the place to focus on for most benefit from optimisation.

ATP will be an enabler for improved line capacity and opportunities to continuously control the speed of trains from a centralised control system. You could then get more consistent and predictable operation. From that point of view, a continuous and comprehensive ATP system is also the way to go for the best potential benefits.

So basically, the important thing is capacity.

Qu 8. What ATP developments do you believe would be most likely to encourage the wider / quicker application of comprehensive ATP systems on UK main line railways?

A ATP system that also delivered operational benefits would be easier to justify. Conversely, if it degrades the performance of the railway by, for example, intervening too early, it would be less justifiable. I think that it would be easiest to justify continuous ATP systems on that basis in the UK.

Qu 9. As an additional point to this, how do you think the trade off between safety and operational flexibility can best be managed?

I think that continuous ATP would provide both operational flexibility and safety, so there would be no conflict!

A lesser system that only does spot checks, like Chiltern ATP can intervene too early because it only looks occasionally and makes worse case assumptions based on that. Spot ATP can have a degrading effect on operational flexibility, but a continuous system would not. Unfortunately, people look at initial costs rather than life cycle cost.

I am not sure about how to draw the line between the level of safety and how much flexibility you want. I will need to think about that further. This would be a particular issue when it comes to degraded performance levels. With compromise systems, this gives problems, but I think that the comprehensive ATP system approach would be safer than current practice and therefore presents less of a problem.

Qu 10. What information do you think:

A. Is required by a driver in order to perform the tasks asked of him

Assuming under ATP.
Maximum allowable speed before ATP kicks in. That is a curve below the emergency intervention curve. The ATP emergency brake envelope is no use to him. The important speed is the one that he has to drive to.

B. Would be beneficial to a driver to assist him in the performance of his tasks

When to leave a station – timetable information.
Maybe which platform he is approaching at a station, to help improve his driving performance. The mental preparation in advance may help improve efficiency.
Also actions to keep the driver alert. So giving him a choice of actions (only one of which will actually be allowed by the system), such as pressing the door open button for the correct side. I.e. protect against his mistakes, but still let him take actions and make decisions.

C. Would be detrimental to a drivers ability to perform the tasks required of him

This is not something that I have given much thought to.

Qu 11. How do you think that your answers to question 10 would change if the system included ATO?

Maybe information to drive off peak.
Some activity such as opening and closing doors.
Maybe it would be useful to have the driver on the platform rather than the train? He is most useful managing the train stops. He does not really need any other role under ATO.

Interview by e-mail with:

Helmut Uebel FIRSE (Retired).

Question Raised by D. Woodland on 28/10/2003. Response given by H Uebel on 29/10/2003.

Qu 1. I found a very interesting paper that you wrote for Signal & Dracht in 1998, 'Durchsatz von Strecken und Stationen (Line and Station capacity', which refers to two examples. One for a passenger train (using a 25m safety margin for speeds of up to 400km/h) and the other for a freight vehicle (using a 100m safety margin at speeds of up to 200km/h). I would be interested to know whether these were references to real applications, a projection of future possibilities or just a hypothetical (not realistic) value to simplify the example.

It is long ago since I wrote this Paper in S+D. I had difficulties in finding the text again, but finally I got it. Now to your question concerning safety distance. There are three aspects to it:

1. It was used in the past primarily for mechanical problems such as slippery rails, partial defects of brakes and so on. This is a decision of the mechanical engineers and cannot be changed by any signalling system.
2. Non-compliance of the driver with speed, excessive speed (braking curve or stopping point). This can be reduced to a great deal by continuous supervision (ATP) and/or automatic operation.
3. Errors in positioning and/or speed measurement. The position error is relatively new and only relevant for Transmission Based Systems.

In practical terms, we use a 25m safety distance in our SELTRAC urban automatic systems with the approval of the respective safety assessors, up to now no problem reported.

In main line systems in Germany, we have a minimum safety distance of 50m in blocks. In reality, the safety distance is much longer, as we assume, that the preceding train has a maximum length of 750m, the average bloch length being 1500m, which gives an average safety distance of 750m in the case of two trains following each other with the same speed.

In this paper, the safety distance of 25m was chosen to demonstrate the extremes; technically it is possible based on the widely installed LZB system.

I left the selection of operating parameters open, this is not only a decision of us as signal engineers. The aim was really to show that there are possibilities to increase the train density by means of "software", not by concrete. There is always the problem in Germany that the railways tell the polititians that the railways are almost at the limit of their capacity, which is not true.

Interview by e-mail with:

Mark Williams (Senior Consultant, Mental Modelling and Intelligent Interfaces Group, BAE Systems Advanced Technology Centre).

Questions Raised by D. Woodland on 10/06/2002. Responses given by M Williams on 12/06/2002 and 13/06/2002.

Qu 1. In order to provide the context within which your subsequent answers should be viewed, could you please provide an outline of your career experience?

My main specialisation is in hearing conservation and noise, but I have been able to look a few things up and and talk to people here who have expertise in human response to vibration and modelling of crashes in aerospace and automotive applications.

Qu 2. I am currently investigating is the physiological effects of the acceleration and jerk (rate of change in acceleration)experienced by passengers. When considering this issue, it occurred to me that passengers on aircraft must be subjected to far higher acceleration and jerk rates than passengers on trains! I would, therefore be very interested to find out any details of what rates do actually occur, and under what circumstances?

I have not been able to find any information on jerk, but there are various regulatory instruments and practices in relation to accelerations and loading. The US Federal Aviation Authority have FAR (Federal Aviation Regulations) and there are also JARs produced by the Joint Aviation Authorities in Europe (of which the UK CAA is a member). The relevant regulations seem to be FAR-23, FAR-25 and 135 and the JAR-23 and 25. These divisions reflect slightly different categories of transport (based on maximum take-off weight) The main point is that, in normal operations, maximum design "g" loadings are "quite" low and, from an aerodynamic/structural point of view, the g rating need not be greater than 3.8 (4.4 for utility aircraft) FAR-23. Negative "g" design load limits are around half these values. These would represent

quite violent manoeuvres which one would not normally expect a pilot to have to carry out. From the point of view of aircraft design there appears to be a rule of thumb that aircraft should basically be expected to fly in a straight-line with "little" perpendicular acceleration as this is not acceptable for passenger comfort (J Roksam and Chuan-Tau E Lau "Airplane aerodynamics and performance", DAR Corporation, 1997 [http://www.darcorp.com]). In the anticipation of turbulence, passengers are usually required to remain seated and with belts in place. Pilots would generally aim to keep passenger comfort as high as possible and so any untoward acceleration would be expected to be avoided unless there was a good reason for it. A return to steady level flight would be attained as soon as possible. Experience of flying as a passenger on many commercial flights suggests that usually there is very little such untoward acceleration. "Buffet", which can arise in a number of ways, can make an aeroplane difficult to control or uncomfortable for passengers. I am not sure how the figures are arrived at but Roksam cites the US Navy standard AS-5263 "Guidelines for the preparation of standard aircraft characteristic charts and performance data" as supplying a definition of "intolerable" buffet (from a military point of view). US Mil-Standard MIL-A 8861 (ASG) provides a table of design limit load factors ranging from 8.5g for fighter aircraft to 2.00g for heavy bombers. Crash experts tell me that 15g (Gx - forward or backward) are thought survivable, so seat manufacturers have to work to a standard which means that a seat must be tested to remain in place, with suitable dummy loading, under such deceleration. Sanders and McCormick (Human Factors in Engineering and Design, 1992, McGraw-Hill) illustrate some of the effects of "g" forward or back. Up to around 3g, reference is simply to "abdominal pressure" rather than other effects like "pain" or loss of peripheral vision. During take off and landing, the accelerations involved are moderate (especially on something like a 747 weighing many hundreds of tons). Thrust to weight ratios are less than 0.5. This is why runways are long! Without reverse thrust, braking decelerations appear to be of the order of 1.5 ms-2. It seems to me that the general assumptions are that one avoids any unnecessary accelerations. In the normal course of operations that is entirely possible but in emergencies high accelerations may occur but are unavoidable and necessary for preservation of life. Limitations are then likely to be airframe/mechanical not human. Comfort considerations are more related to noise and vibration and organisations like the SAE (there is a sound and vibration CD ROM) can provide papers on such things (there may also be some papers related to seat belts).
There was a study, which given your background you may already be familiar with, done at ISVR, author Paddam, for London Underground on this sort of thing. It included examined how people hold onto the handgrips, and so forth, and using this to determine maximum accelerations. I don't have the detailed reference (and I am not sure of the spelling - my colleague who mentioned it yesterday is not in today). But this may be enough for you to go on or not!

At face value, the figures you quoted of 3g leading to 'abdominal pressure' but no injury seems surprisingly high. I suspect that such high deceleration has so little effect because aircraft passengers wear seatbelts, have no loose luggage and are travelling along a fairly smooth / object free runway?

I quite agree. The figure I mention from McCormick (3g)would represent, probably, data taken from centrifuge measurements with people strapped into seats. Subjects would no doubt report what they felt and physiological measures may have been taken also. In Salvendy (see below) the narrative actually suggests that at 2-3g it becomes nearly impossible to move, so one must take these simple labels in context! At higher g (> 10) levels, this may represent data taken from crashes and calculating maximum levels experienced etc. I haven't had time to pursue this thoroughly, but I just think that there are probably different design considerations in aerospace. That is, it is unlikely that maximum levels would be specified for unbelted passengers because in circumstances where there was turbulence etc. there is a threat to the aircraft (a pilot may know there is bad weather ahead, but won't necessarily be able to predict what level of threat it presents to the aircraft, therefore the procedure would be to avoid it.) Also an aircraft tends to make just one important, short start and one important, short stop in its journey. Seatbelting is important because of the desire to avoid the "second collision", as it is called. Salvendy (Wiley) "Handbook for Human Factors and Ergonomics" also contains a chapter on the effects of "g" (micro and macro-gravity) should you wish to pursue another reference. I have looked through an ex-colleague's papers on crashes but on a cursory glance haven't found any mention of "g" levels. There are plenty of references to accident indices, however! From these papers it seems that one could expect accelerations of parts of the body in the order of 20g in automotive crashes

Telephone interview with:

Nick Wright IEng, MIIE(elec), AMIRSE (Trainborne Systems Manager, First).

The author's questions were provided in advance by e-mail. Answers from Nick Wright are based on notes taken during a telephone conversation on 23 /07/2002 and during a group discussion on 4/10/02

Qu 1. It is my understanding that the BR ATP schemes provide:

- **Speed supervision against the maximum speed permitted for the current location of a train. An audible warble and visual warning to be given to the driver if this is exceeded by 3mph or more and an automatic full service brake application if it is exceeded by 6mph or more;**
- **Braking supervision to ensure that the train will slow down sufficiently to achieve a target some distance ahead. To be achieved by calculating a target speed curve and then treating this in the same manner as a maximum permitted speed (an audible warble and visual warning being given to the driver if it is exceeded by 3mph or more and an automatic full service brake application if it is exceeded by 6mph or more);**
- **An indication to the driver of the current or target speed limit. Any new target being displayed in time to give the driver the equivalent of 3 seconds running time at the current line speed before a warning for under braking would be required at the current actual speed.**

I would appreciate knowing from you:

a. **Whether my understanding is correct;**

In general terms, yes it is.

b. **How suitable this approach has proved in operation and;**

One of the biggest problem has been with in-fill. If the ATP sights the signal before it changes aspect, without infill it then forces the train to crawl up to the signal (possibly half a mile or more away). Before infill was installed extensively, this caused a huge impact on performance. Infill has now been installed at all performance critical signals, but the problem still occurs too frequently.

c. **Whether / how you would like to change it in the light of operational experience.**

No specific changes in the areas outlined above. What would be useful is a change in the way that ESR's are dealt with – in the implementation, supervision and display to the driver. At the moment ESR's are applied by installing a data plug in the trackside equipment. When the Trainborne ATP equipment first detects data from an ESR plug, an audible warning sounds, which the driver must acknowledge after approximately 2 seconds. If he does not, the service brake is applied and the train is brought to a stand. The letters 'ESR' are also displayed on the alphanumeric display on the drivers MMI, reminding him that the train is within an ESR area. The audible warning only sounds following the first detection of ESR data.

The ESR plugs are stored in the trackside location cases and contain all of the data required for full supervision to the PSR speeds and aspect status information for that location. As it contains the normal PSR data, the ATP continues to provide supervision against that. The ATP does not supervise the ESR. Compliance is left to the driver.

The in-cab display is reset when the Trainborne equipment receives data from the trackside that the driver is clear of the ESR area . The Trainborne system then returns to full supervision. If another ESR transmission occurs again after that, the audible warning will occur again.

TSR plugs are programmed especially to contain the characteristics of a restriction and can, therefore, enable full supervision. However, it takes time to get a plug programmed with the correct data.

The new class 180 trains that have just been introduced have shorter system delays within the Train braking systems. As a result, the ESR warning causes almost immediate brake application which can be in excess of 16%g! A modification is currently being investigated to re-introduce the 2 second delay.

d. **I am particularly interested in whether the use of line speed, rather than actual speed, when determining the initial indication to a driver of an approaching change in target speed has caused any operational difficulties.**

No difficulties due to this have been observed.

Qu 2. It is my understanding that the BR ATP schemes have no provision for adjusting the braking calculations used in supervision in the case of known poor adhesion.

I would appreciate knowing from you:

a. **Whether my understanding is correct;**

Yes it is.

b. **Whether you have any poor adhesion detection and compensation system that assists the driver in managing poor adhesion?**

No they do not.

What is done to mitigate the risks associated with poor adhesion (apart from the current practices such as vegetation control, etc)?

Adhesion is left to the driver to manage. The ATP is only a background supervision system, and it does not interact with the traction to deal with the problem of poor adhesion. Although, it can provide a degree of compensation to enable operation to continue through the effected area.

c. **How suitable has this approach proved in operation / have any difficulties been encountered?**

Initially, the odometry correction was not very good, leading to beacons being passed too early "sliding" or too late "slipping" (as the train saw things). This initially resulted in non-fatal errors being displayed to the driver, ultimately in a fatal ATP fault being declared and the brakes being put on. It also caused problems with oscillations between slip/slide and traction/braking that led to problems with loss of air in the brake system and excessive wheel damage through flats. Errors such as these still occur sometimes (but not often, due to improved software compensation). Obviously, putting the brakes on actually makes the situation worse.

It got so bad in 1997 autumn season that the whole ATP system was switched off to avoid further delays.

One of the main problems with all of this was severe loss of driver confidence.

TPWS is also beginning to lose driver confidence, particularly on approach to buffer stops. The TPWS transmissions were not very effective at low speeds, leading to unnecessary trips far too frequently, although a change to the trackside loops has improved the situation.

d. **The ERTMS system requirements specification requires provision of a low adhesion input (either operated by the driver or automatically by the trackside). This is intended to reduce the assumed braking rate in the calculation of supervision envelopes to 70%. The TCS project are attempting to assess the need for / usefulness of this feature.**

I think that this sounds like a useful idea if worked by switchable balises, but too constrained. It would be unlikely that a single figure could be optimal for different types of trains and adhesion conditions.

I see adhesion as an availability / reliability issue with safety implications, rather than as a safety problem.

Qu 3. **It is my understanding that once a movement authority has been given by a beacon transmission on the GWML BR ATP scheme, it is assumed to remain valid until the next transmission has been received. I would appreciate knowing from you:**

a. **Whether my understanding is correct;**

That is correct. If the change of aspect occurs before a balise is reached, passing the balise at what is now too high a speed causes an immediate brake application. If it happens after passing a balise, the ATP has no reaction until after the next balise is passed. In-fill loops only provide minimal safety benefit and do not force revised slower speed supervision.

b. **Have any difficulties been encountered in operation on GWML due to route revocation.**

I am not aware of any.

c. **There is concern on the WCML TCS project that if a route is taken back in front of a train after a movement authority has been issued to it, and the train is then delayed (perhaps at a platform) for a sufficient time to allow release of the signalling locking of the route ahead, the ATP system may end up believing that it has an authorised route (and indicate this to the driver), when it in fact does not. This would undermine the safety supervision and, due to the in-cab indication, could increase the risk of driver error over the same scenario without ATP. We are currently trying to assess how much of a problem this scenario represents in order to determine whether it is worth trying to find a technical solution.**

The driver should deal with this situation in accordance with normal driving practice. He/she is not supposed to take the ATP's word for what authority there is. That comes from the signals. This situation is also rare and route revocation is not normally encountered.

Qu 4. Data entry

a. What parameters does the driver have to enter before starting a journey?

On opening the desk, the driver is asked if it is a standard consist. If not, he is asked to enter the parameters. Once they have all been entered, he is asked to confirm that they are correct.

Parameters such as train length are entered as number of vehicles. All GWML passenger vehicles generally have standard vehicle lengths. The standard brake rate is also a data plug. All the driver changes is the state of equipment. Units isolated, etc.

b. How does BR-ATP cope with freight trains in terms of braking model/data entry (compromise between performance and safety)?

We do not have any fitted freight trains. On a freight train, you may need to do something else (since the brake rates and vehicle lengths may vary more widely). The original system was designed to support differential speed control, although the message functionality was removed in order to support other requirements.

Qu 5. The ATP Evaluation and Steering Group 'Formal Evaluation of the BRB Automatic Train Protection Pilot Schemes Report' recommends that future systems should also consider suppressing TPWS. Is there any practical experience to show that such an approach is of significant benefit?

Drivers use the AWS bell and horn to help them locate where they are, particularly in adverse weather conditions. BR-ATP and TPWS do not have this facility. ERTMS may have to some extent. Such location assistance is one of the few reasons that AWS is retained once a reliable ATP system is in operation.

On the other hand, duplicated systems result in poorer overall reliability and greater overall delays. It is hoped that suppression of AWS and TPWS would overcome some of this.

Drivers also complain about the overload caused by all of the duplicate systems: ATP, AWS, TPWS, DSD, NRN, ... It could be considered that there is already too much for the drivers to deal with. Suppression would, therefore, significantly reduce driver distraction, leaving them free to spend more time on actually driving.

If future ATP systems could provide the location awareness that is currently given by the AWS bell/horn, AWS and TPWS should be suppressed on the grounds of Human Factors.

Qu 6. Release Speed

a. How is the driver informed of a transition into release speed supervision, as the train approaches a beacon protecting a signal that has cleared, in a manner that mitigates the risks associated with driver distraction and head down driving tendency at a critical time?

The driver's speedometer has green LED indicators around it for ATP speed targets. There is also a second ring of yellow LED's at lower speeds. These indicate release speed. When activated, they light up so that the driver knows what the maximum allowable release speed is. This is achieved through data received from the trackside beacons. When in-fill is fitted, then release speeds do not apply.

b. What variables is the release speed calculation based on?

A combination of train and track characteristics. Braking rate, overlap length and gradients

c. Do timed overlaps affect how release speed is handled by the system? If so, how?

I am not aware of there being any timed overlaps on GWML.

The biggest problem experienced is at Reading, where two of the platform starter signals have no overlap available. They have a very low release speed. However, since drivers do not approach red signals closely anyway, this does not cause too much of a problem.

Qu 7. What factors does the BR-ATP braking model take into account e.g. does it account for the mass distribution, mass in general, rotating mass, length of train, brake build-up time, gradient, etc? If so, could you provide any details on tolerance, range, etc. required of any associated input data?

I do not have that information. You could try contacting the Alstom project manager:

Peter.van'twesteinde@transport.alstom.com

Qu 8. Has BR-ATP experienced excessive braking distances after a series of brake application and release cycles (either by the driver or ATP intervention) over a short period of time, as a result of progressively increasing brake build-up time?

Yes, particularly when the oscillating effect was occurring during slip/slide. That has now been resolved by software upgrades.

Qu 9. Is it correct to presume that all BR-ATP data entry, usage and display is in imperial units (i.e. mph for speed display and trackside speed restrictions, miles for train location, etc.)?

The algorithms are in metric, display in imperial. The analysis software shows both. (Speak to Peter Van'twesteinde for details).

Qu 10. Is there any BR-ATP protection provided at shut down cabs (e.g. rollaway protection when no cab is open)?

It is important to first consider the opening of cabs. About 4 or 5 times a month, the self-test fails on one of the units without any obvious reason (there are about 70 in use a day). We believe this is due to EMC problems, but have never been able to track it down. Initially, the self-test took 5 minutes and the train had to be withdrawn from service if it failed. To help resolve this problem, we now have a reduced self-test on opening that only lasts 90 seconds. If it fails, the self-test can be repeated and the train is only withdrawn from service if it fails again. The full 5-minute test is only done in the depot.

When shut down, the ATP does not provide any protection functions . Rollaway protection when the ATP is shut down is provided by the parking brake.

Qu 11. Does the BR-ATP system take into account the existence of WSP in terms of assumed brake rate? That is, is the nominal brake rate influenced by the WSP's existence?

Not directly, hence the oscillations between brake application and release that were initially experienced. That has now been overcome by software changes.

Qu 12. Is there a legal data recorder onboard the BR-ATP fleet?

- **If so, what does it record?**

Everything that happens is recorded. The biggest user of data is the raw tachometer input (that is a square wave and takes a lot of space). The processed speed calculation is also recorded, but takes a lot less space. The data memory is 2MB.

How long does it record before the memory is overwritten with new data?

The minimum time to rerecord so far has been about 48 hours.

Qu 13. Is BR-ATP aware of/does it use coupling status information?

Vehicle length and consist information is confirmed as part of the ATP Data Entry procedure only. On 180's all coupling information is by data entry. There is no automatic ATP coupling detection. The driver receives a prompt 'is standard consist applicable' on start up. If the answer is no, each piece of individual information has to be entered. Once that has been done, confirmation that the data is correct is then asked for.

Qu 14. What speed thresholds and time constraints are used to determine when a train is at standstill?

This is not provided.

Qu 15. Does BR-ATP request different braking types according to different circumstances (e.g. service and emergency brakes) or is one type used for all scenarios?

There are two types of request, but only one brake. The difference is in how quickly the braking system responds to the request.

Qu 16. Output Status Feedback

a. Does BR-ATP request traction inhibit (and does it get feedback of this status)?

Yes, it does, via the brake demand relay.

b. Does BR-ATP obtain brake status feedback? If so, from where (e.g. via a relay, brake pipe pressure, etc)?

Not directly. The ATP looks for speed reduction. Without it, the EB is also requested. The ATP does look to check that the brake demand relay armature has moved following a demand, but does not monitor the front / back contacts themselves.

Qu 17. What is the composition and recommended frequency of:

- **Power-up test (e.g. are all trainborne ATP peripherals tested)?**
- **Daily test (e.g. is ATP brake/traction inhibit output tested)?**

On opening the cab desk, which normally occurs at the start of each journey, the ATP performs a self test sequence which ensures all peripherals are tested by applying and releasing them and asking for driver confirmation that the in-cab instruments show this happened correctly. E.g. that the brake pressure has dropped off and come back up again, etc.

If a fatal fault occurs, this is displayed on the three digit alpha numeric display as 'F12'. The brakes come on, and the train has to be removed from service.

If a recoverable fault occurs, this is displayed as an 'E' fault. The driver has to log the occurrence of this, but can continue the journey. Data can be downloaded from the VOBC at the end of the journey to help locate and fix the cause of the error.

If one of the processor channels fails, the driver is not informed. The ATP continues on a 2 out of 2 rather than 2 out of 3 basis. This has caused a lot of problems by masking failures. It would be better to inform the driver of the first channel failure, so that he can report it and maintenance can be arranged before a second channel fails. When the second channel fails, the train has to be withdrawn from service.

Problems have occurred with loose wires simulating a desk shutting down and re-opening during operation. This causes the self-test to run through, including application of the brakes. This has occurred far too commonly (around once a month) and a modification is being looked at to improve the integrity of the wiring.

A very lengthy and comprehensive set of tests can be carried out in the depot by means of a very cumbersome piece of test equipment. It takes about an hour and is difficult to do.

Qu 18. I note from the ATP Evaluation and Steering Group 'Formal Evaluation of the BR Automatic Train Protection Pilot Schemes Report' that the evidence from (lack of) SPADS during the evaluation period is not statistically significant and does not yet demonstrate an improvement in safety. This raises a number of questions:

a. Is it possible to draw any conclusions from evidence about incidents when BR-ATP has given warnings to the driver or initiated a brake application, that the risk of SPADS has been positively reduced?

SPADs have gone down, but it is impossible to tell if that has been due to the ATP or other changes. It is also impossible to tell what the risks were at the time of a brake intervention or warning without detailed analysis of the data recorder outputs. To date only 6 of the GWML trains have been fitted with recorders, so it has not been possible to tell.

Since warnings will go off repeatedly on approach to critical locations, such as a terminal platform's buffer stops, is it acceptable to assume that a warning would have been unsafe? I don't think so, as long as intervention is avoided. The driver may need to drive on the warning curve in order to get optimal performance in that area.

b. Are the numbers and types of ATP intervention logged and analysed to determine which have been genuinely valuable to the safe operation of the train?

You can get a comprehensive download.

c. Given that there must be/have been considerable pressures to adjust ATP operation to minimise the impact on the driver and train performance, how has the project maintained the focus on ensuring that a useful improvement in safety is actually still achieved, rather than just implementing a system with the least operational dis-benefits?

Today approximately 99% of drivers see the ATP as a beneficial driving aid. Now it is working reliably , they seem to like it. Commercially it hurts, by impacting capacity (particularly when failures occur). That is something you have to live with.

Qu 19. I understand that you now operate a no-ATP no-go policy. What happens if a train's ATP fails in service?

The driver must stop immediately and report the failure to Railtrack. He calls control, who will advise on the course of action to take. This will be:

- Take out of service;
- Pick up a second man at the next station;
- Turn the train, so that the working ATP on the other cab can be used (train's with a failed trailing end ATP are not permitted to enter Paddington station).

The train cannot just continue on the journey. Trains can also not enter an ATP area with failed ATP on the leading end.

Qu 20. Do you have a separate speedometer to the ATP one, for use when the ATP fails?

At the moment most trains do. However, there is an ongoing programme to remove them. The ATP is now considered reliable enough to avoid the need for a separate speedometer.

APPENDIX B - TRAIN CONTROL SYSTEM AS A CLOSED LOOP CONTROL SYSTEM

INTRODUCTION

During the course of his background research, the author developed four representations of train control systems as closed loop control systems. These can be seen in Figures 3.2 (Basic Human-Machine Train Control System), 3.3 (Advanced Human-Machine Train Control System), 4.6 (Train Control System With Automatically Closed Safety Loop) and 4.7 (Fully Automatic Closed-Loop Train Control System).

Subsequent to the development of these representations, several equivalent representations of railway control systems have come to the author's attention within published literature. The purpose of this appendix is, therefore, to consider these representations and compare them with the authors own.

DR COLIN GOODMAN

Within a paper on ATO, ATP and ATC, Dr Goodman gave the representation of an ATO system shown in Figure B.1.

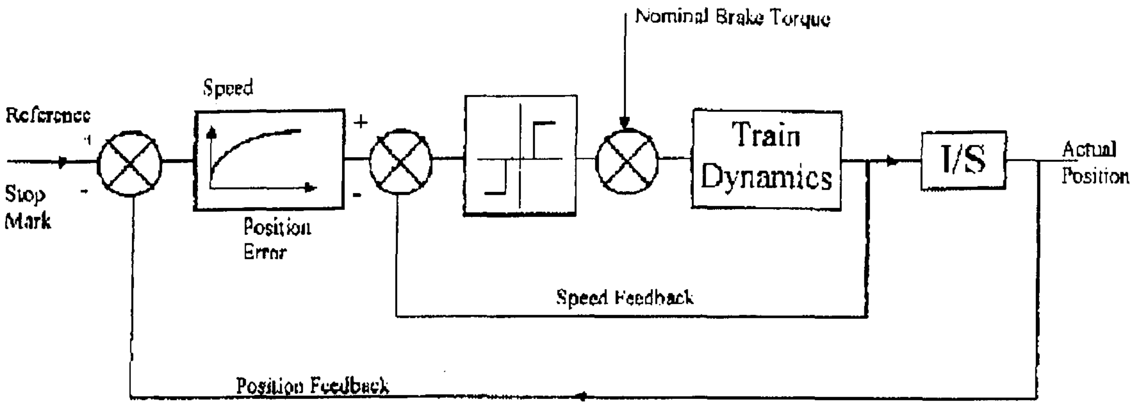


Figure B.1: Dr Goodman's ATO Closed Loop Control System (Goodman 1999, p7)

This representation is equivalent to that given in Figure 4.7 (Fully Automatic Closed-Loop Train Control System), but limited to consideration of the trainborne ATO elements of the control system. Hence, the input to the control loop is seen to be a reference stop mark, rather than the line condition ahead. This is then compared with actual position in order to generate a speed target. That is in turn compared against speed feedback in order to determine the brake / traction demand required from the trains dynamics. No provision is made within this representation for intervention by a safety loop, should the speed demanded by the position error actually be in excess of line or train maximum safe speeds. It is possible that Dr Goodman intended the input function 'reference stop mark' to include the safety constraints that would be required for ATP type speed supervision, but this can only be implied by an understanding of the system that is being represented and is in no way made clear within the representation given. In the opinion of the author, this representation is therefore incomplete.

There is one element within Dr Goodman's representation that is not included within Figure 4.7. That is the additional comparison of brake / traction torque demand and 'Nominal Brake Torque' used to derive the actual control of train dynamics. This element of the control loop is not explained within Dr Goodman's paper, but it is assumed that it has been included within the representation in order to highlight the need to translate demand for brake / traction torque into actual application of the brake / traction systems. On reflection, it is the author's opinion that this function is a part of the brake control sub-system and hence a level of detail below that shown in Figure 4.7 (which is intended to represent the train control system as a whole). It would in fact be a part of the 'controller' transfer function. It is not, therefore, considered appropriate to explicitly include it within Figure 4.7.

ROGER SHORT

Within a paper on Fundamentals of Signalling and Train Control, Mr Short gave a number of representations of different types of railway control system, corresponding to the author's own representations to a much greater degree than Dr Goodman's.

Mr Short begins by considering the feedback control system of figure B.2, which he describes as a representation of a conventional signalling system (the author's view of which can be seen in Figure 3.4)

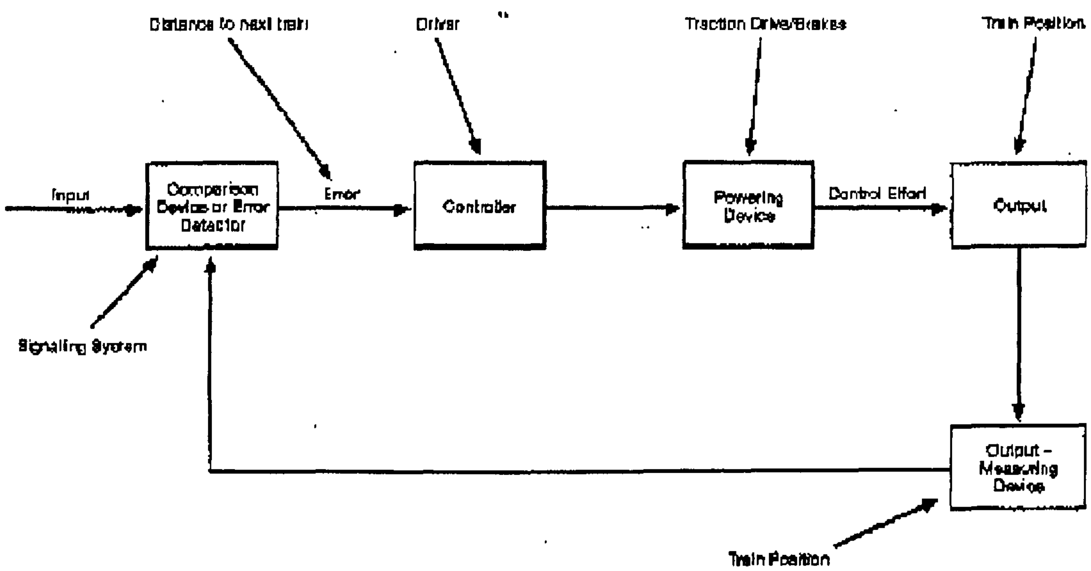
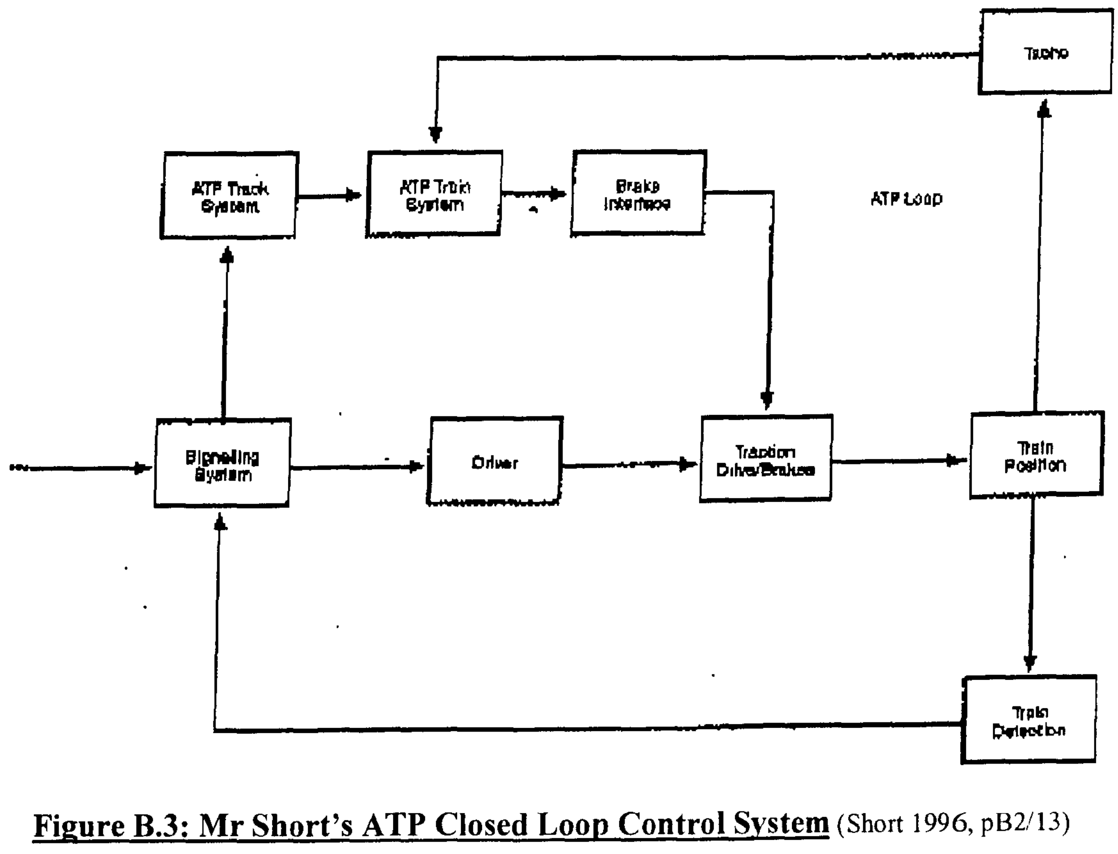


Figure B.2: Mr Short's Feedback Control System (Short 1996, pB2/9)

All of the elements within Mr Short's representation are included within Figure 3.4. However, they are not all represented in the same way, or as having the same functions. In particular, Mr Short denotes the signalling system as providing the error measurement by which the driver controls the train's brakes and traction. In contrast to this, the author sees the signalling system as providing target information (which will largely be based on the position of other trains) in the form of a signal aspect that can subsequently be used by the driver to determine the error measurement. The driver would be expected to do this by comparing the signal aspects with the train's speed and his/her assessment of the current train location. It is the author's opinion that the representation given by Mr Short does not, therefore, fully highlight the role played by the human driver within the control process. This point can also be observed from the fact that Mr Short does not differentiate between the parts of the control loop that are achieved by technological solutions and the parts that are achieved through manual activity.

Whilst considering Mr Shorts representation of the signalling based closed loop control system, the author observed that the comparison of signal aspect and speed shown in Figures 3.4 and 4.6 as a function of the driver's eye and mind would also require consideration of the trains position, which had not previously been included within the figures. It could be argued that an additional feedback loop of position information is not required since it is in practice achieved by the drivers observation of signals and other lineside markers (the observation of which he/she then converted into positional information based upon his/her route knowledge). The feedback of some positional information is therefore implicit within driver's observation of signal aspects as represented in Figures 3.4 and 4.6. However, Signals are not the only source of information used by the driver to determine the train's location and judge distances. The Author therefore decided to include this additional feedback loop within the figures.

The second representation of a railway control system given by Mr Short is for an ATP system, shown in Figure B.3, which is comparable with the author's representation of a train control system with automatically closed safety loop in Figure 4.6.



<u>Figure B.3: Mr Short's ATP Closed Loop Control System</u> (Short 1996, pB2/13)

Once again, all of the elements within Mr Shorts representation are included within Figure 4.6. The same differences occur as have already been discussed for Figures 3.3 and B.2. The arguments will not, therefore, be repeated. In all other respects, the two representations appear to the author to be compatible (with the obvious exceptions of terminology and diagrammatic representation). The author does not, therefore, have any further comments to raise on Mr Short's representation.

An additional question could be raised, however, about Figure 4.6. The figure does not show any position feedback to the ATP system. Such feedback may actually be required by some types of ATP system (for example a comprehensive distance to go based ATP system), but would not be necessary for other types of ATP (such as a comprehensive speed code based system). In order to simplify the diagram, the author has therefore excluded such feedback from the representation shown in Figure 4.6.

The final representation given by Mr Short, shown in Figure B.4, is for an ATC system, including both ATP and ATO functions. This is comparable with the author's representation of a fully automatic closed-loop train control system in Figure 4.7.9

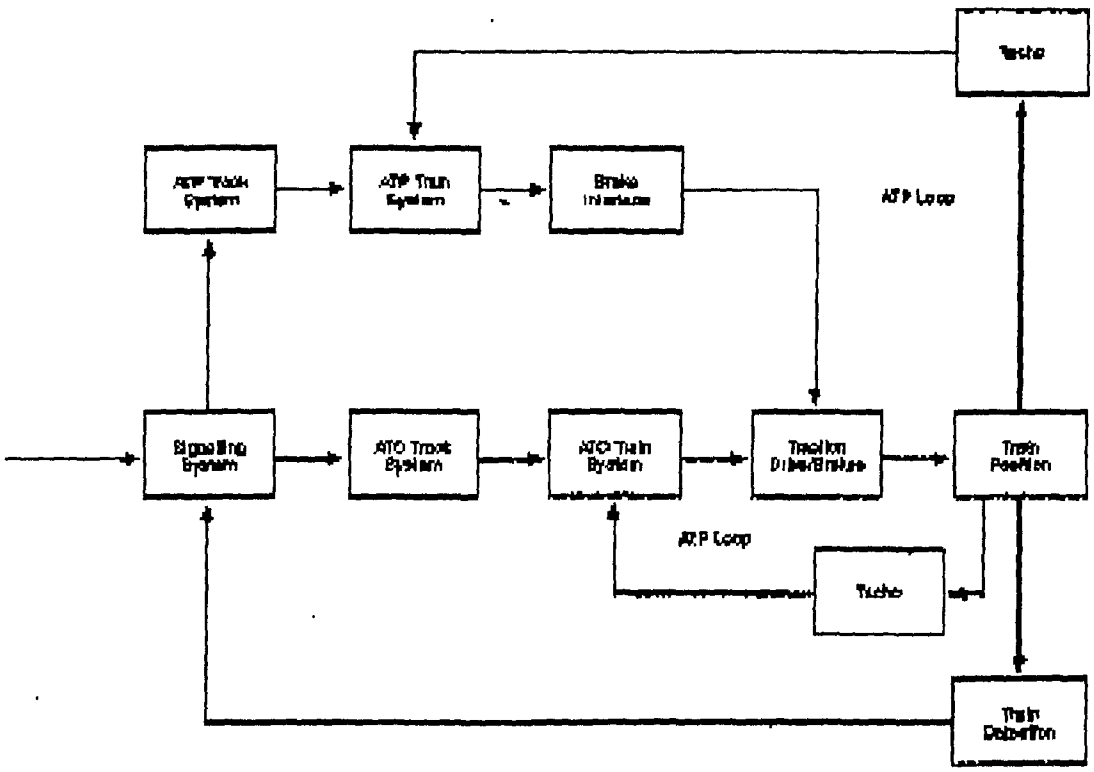


Figure B.4: Mr Short's ATC Closed Loop Control System (Short 1996, pB2/14)

Here, there are a number of minor differences between Mr Short and the author's representations, which are none-the-less worthy of note.

The author uses a 'controller' to interpret the ATO system error output and to then control the trains brake / traction systems on the basis of that interpretation, where as Mr Short shows a direct interface between the ATO and traction drive / brake systems. On reflection, it is the author's opinion that this is not a material difference, but one of functional allocation only.

Of more interest is the fact that Mr Short's representation does not show any direct feedback of the train's position to the ATO or ATP systems (only a direct feedback of speed and indirect feedback of location via the signalling system are shown). As already discussed above, this would be a valid representation for a speed code based system (such as the FS2000 system used on LUL's Central Line). In this case, however, the author has chosen to represent a distance to go type system, and has therefore included feedback of position to both the ATO and ATP system.

The author has not observed anything within this representation that would imply the need for changes to his own representation (as shown in figure 4.7).

MIKE CHANDLER

Within a paper about the Central Line ATO Algorithm Mr Chandler provided the control loop representation of figure B.5. This representation is equivalent to that given in Figure 4.7 (Fully Automatic Closed-Loop Train Control System), but limited to consideration of the trainborne ATO elements of the control system. As such, it is very similar to Dr Goodman's control loop that has already been considered. It is, however, more detailed in its representation of the control implementation.

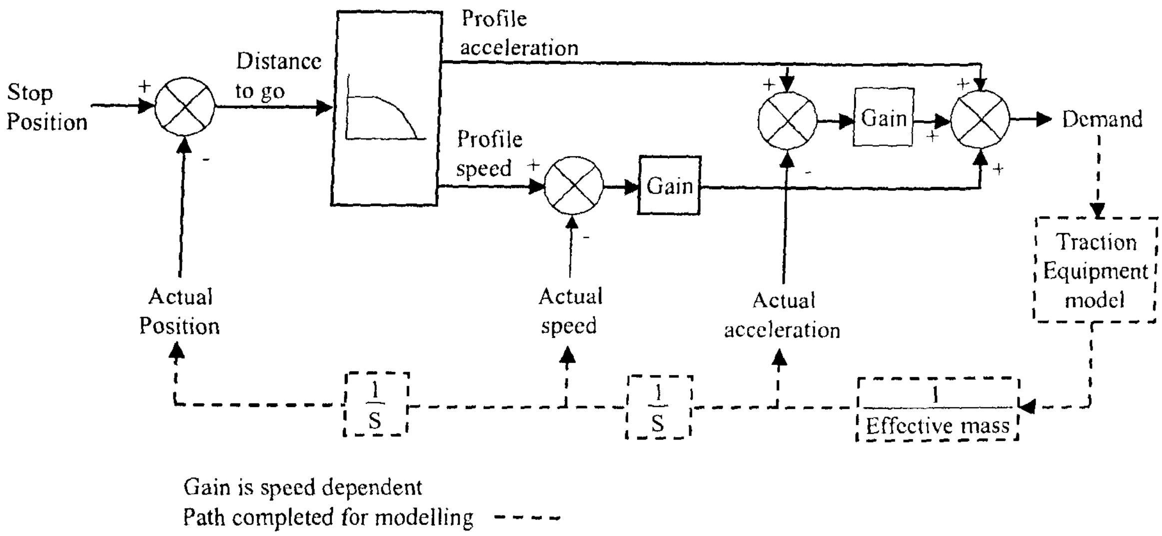


Figure B.5: Mr Chandler's ATO Closed Loop Control System (Chandler 2001, p30)

Any detailed review of Mr Chandlers control loop would involve repetition of the observations already made for Dr Goodman's representation. Therefore, the author doe not propose to provide comment beyond stating his opinion that it does not raise any issues that would require modification of Figure 4.7.

APPENDIX C: AUTOMATIC TRAIN PROTECTION, SYSTEMS AND APPROACHES

C1 INTRODUCTION

In section 4.3.2, Automatic Train Protection (ATP) was defined as 'a system automatically enforcing compliance with **or** observation of **some** or all speed restrictions **or** movement authorities'. The interpretation that can be applied to meeting this definition has lead to a plethora of systems that could be described as forms of ATP. It is the author's intent to consider a comprehensive (but by no means exhaustive) selection of these within this chapter, in order to highlight the variety of ways in which train movements can be protected automatically, along with some of the limitations associated with the different approaches.

In order to provide as complete a picture as possible of the development of ATP systems it is the author's belief that recognition should be given to the role played by manually activated systems. Therefore, this chapter will commence with a consideration of warning systems, before moving on to a discussion of different ATP systems and approaches.

C2 WARNING SYSTEMS

Train protection systems were developed out of a realisation that relaying control information to a train by use of visual signals carried with it inherent risks that could be reduced. The first train protection systems were also technically the simplest, in the form of warning systems. These were not intended to introduce a reliance on safety equipment, but rather to assist the driver who retained full responsibility for the safety of his/her train. In section 4.3.2, a warning system was defined as 'a system assisting observation of movement authorities, based upon manual activation'.

An example of such a system was used for repeating semaphore signal aspects to the driver in the early days of railways. Drivers were generally expected to have sufficient route knowledge to know where they were at all times and, thus, when they needed to keep a look out on the approach to a signal. However, in conditions of poor visibility, such as fog or falling snow, it was recognised that they needed some assistance. A 'fogman' (usually from the permanent way staff) was therefore stationed at each signal during poor visibility. If the signal was clear, the fogman displayed a hand lamp to repeat the signal aspect. If the signal was at danger or caution, he placed a detonator on the rail. The sound of the detonation then carried the signal aspect to the driver in his cab (Hall[(1)] 2000, p90; Kichenside et al. 1998, pp156-7). Unfortunately, this system suffered from a dependence on the fogman performing his/her duties correctly in an unpleasant and dangerous environment.

Another example a warning type system used on UK railways is a fairly recent development (with installation completed in 1998). This is known as a Driver Reminder Appliance (DRA). The DRA is a device that, when operated by the driver, disengages the traction system and thereby prevents the train from starting until the driver resets it. The system relies entirely on human operation, with drivers instructed to activate it whenever they stop at a red signal. Having activated the system, the requirement to reset it then acts as a reminder to the Driver that he should check that the signal has cleared before departing (HMRI[(1)] 2000). This system was developed in order to address the problem of start away (or in colloquial language 'ding ding and away') type SPADs, which have historically accounted for around 25% of all losses arising from SPAD incidents (Muttram 2001, p2). Unfortunately, this system also has a number of limitation, including:

- Its inherent reliance on the driver remembering to use it before it can act as a reminder for him/her to check the signal aspect;
- The fact that it may only be operated once the train has come to a stand;
- It "suffers from the repetitive use syndrome of AWS and requires rigorous driver management to ensure it is effective" (Fenner 2000, p7).

As a consequence of these limitations, the effects of the DRA system have not been as good as had been hoped. In particular, a number of incidents have occurred where the DRA was not used even though fitted, with experienced drivers being found to be less likely to use the device than their inexperienced colleagues (whose initial training included use of the device) (Muttram 2001, p2). In an attempt to overcome this limitation a new version of the DRA is being developed that will activate automatically, whilst still requiring manual cancellation by the driver (Uff et al. 2001, p94). When this system becomes available, it will be a form of automatic warning system, as described in the next section.

C3 AUTOMATIC WARNING SYSTEMS

As railway technology developed, methods for providing automatic reminders and warnings to a driver became possible, enabling the development of systems to automatically assist observation of movement authorities by the driver. In section 4.3.2, such systems were defined as Automatic Warning Systems.

As early as 1867, a patent was granted in the USE for a system that could sound the whistle and shut off steam (Aldritch 1993, p52). The author knows no further details of this system, but numerous equivalent systems have arisen on different railways since then.

C3.1 NORTH EASTERN MECHANICAL CAB SIGNALLING

The first of these automatic systems in the UK was installed (on a small scale) for the North Eastern Railway (NER) in 1894. It was intended to operate a whistle in the cab of steam engines as they approached a distant or home signal at danger and, on some engines, it also operated the brake (Fenner 2000, p3).

A metal bar with two protruding 'trip' arms at either end was connected to the signal wire. When the signal was cleared the signal wire rotated the bar, lowering the trip arms. If the signal was replaced, or the signal wire broke, a balance weight rotated the bar in the opposite direction to raise the trip arms. The steam engines were fitted with a vertical striker arm, rounded on the end and aligned to collide with the ground mounted trip arm when it was in the raised position. As the trip and striker arms collided, the striker arm lifted and operated a valve, admitting boiler steam into a whistle in the cab. Where the engine was fitted with an air brake, the mechanism used the brake air supply to sound the whistle and, in the process, initiated a brake application. Both the whistle and brake application remained effective until reset by the driver (Kichenside et al. 1998, p157).

This system offered the driver an indication that could assist in determining both his/her position (which was particularly useful in conditions of poor visibility) and the state of the road ahead. However, it suffered from being discontinuous, with the driver able to cancel a danger indication and then forget that he had done so. The system relied upon the same warning being given for both home and distant aspects and also had the potential for undetected wrong side failures if the trip arm broke off. However, it was found to work reliably.

Plans were developed to replace the bar at stop signals with a centrally located trip arm and install a second striker on the locomotives to match. This would have given separate caution and danger indications in the cab. However, when the NER was incorporated into the LNER in 1923 these plans were scrapped and the system was gradually taken out of use.

C3.2 GREAT WESTERN AUTOMATIC TRAIN CONTROL

In 1905, the Great Western Railway began installing a system known as 'Automatic Train Control' (Barwell 1983, p104). This system utilised timber and steel ramps of 40 to 60 feet in length, placed at a slight angle between the running rails (to prevent wearing a groove in the contact shoe). Initially the ramps were installed 440 yards before a distant signal, although this was reduced to 200 yards by BR (Calver 2001, p2). As a train passed this ramp with the signal displaying caution, a plunger located on the underside of the locomotive engaged with the ramp and was forced up. This caused a trainborne circuit to be broken, which, in turn, allowed an electro-pneumatic valve to open, admitting air to the vacuum brake system and gradually applying the brakes. The passage of air entering the brake system operated a siren, which acted as an audible warning to the driver. The driver could then cancel the warning and brake application by lifting a handle in his/her cab that restored current to the controlling circuit. If the distant signal (and by implication, all of the associated stop signals) displayed a clear aspect, the ramp was energised by a d.c. voltage. If a train then passed over it, the plunger would pick up a current from the ramp, which would maintain the control circuit through a second winding on the electro-pneumatic valve and prevent a brake application / warning. The circuit was also arranged so that the flow of current through the plunger operated a relay that controlled the ringing of a bell in the drivers cab, thus providing him/her with an audible 'clear' indication (Barwell 1983, p104; Calver 20001, p2; Dapre 1999, p2). As absence of the electric current indicated a caution aspect, the system was fail safe as long as the ramp and plunger remained mechanically intact (Hall[1] 2000, p157).

On single line track, the polarity of the current energising the ramp could be reversed. This still operated the second winding of the electro-pneumatic valve (avoiding a brake application), but did not operate the bell. This feature was also used in tests during 1946-47, intended to develop the system to respond to a double yellow as well as a single yellow signal. However, this experiment was not taken any further by BR (Calver 2001, p2)

The use of audible warnings associated with both caution and clear aspects gained the attention of the driver wherever he/she was looking and was, therefore, a big improvement on earlier systems (Calver 2001, p2). As well as improving safety, the greater confidence that this system gave to drivers during bad weather helped to reduced delays (Dapre 1999, p2). However, it still suffered from being discontinuous, with the driver able to cancel a danger indication and then forget that he had done so. As there was no visual display to accompany the audible warnings (Hall 1988, p20), the driver had no permanent reminder of the aspect displayed by the last signal passed. The system also required an on-board power supply (provided by a battery) which then needed to be kept charged for the system to work.

Plans were developed to improve the system, and a trial installation including a separate danger indication at stop signals (by use of different polarity) was demonstrated to the Railway Inspectorate, but not developed further (Kichenside et al. 1998, p159-60). Following nationalisation, the Great Western ATC system was modified to include a visual indication to the driver, in the same form as that implemented for the Strowger-Hudd and BR AWS systems (see sections C3.6 and C3.7). Following this modification, the system remained in use on ex-GWR lines until the 1980s (Calver 2001, p1)

C3.3 THE MILLER SYSTEM

A similar system to the Great Western Automatic Train Control came into use in the USA in 1911 and remained in service until 1950. This was the patent Miller system, installed on the Chicago and Eastern Illinois Railroad. It employed a ramp and shoe almost identical to that used on Great Western, but it was installed in the right hand six foot. On the locomotive, the system used a chain and pneumatic piston to mechanically rotate the automatic brake valve to the service position if the driver did not acknowledge a restrictive signal (Calver 2001, p2).

It is claimed that Miller developed his system based on the GWR ATC system, with slight modifications to make it more suitable for the American conditions (Calver 2001, p2). It was found to operate reliably and to be very durable. However, the Miller company ceased trading and the system was then removed when it became due for renewal (Calver 2001, 2).

C3.4 NER ELECTRIC CAB SIGNALLING

Following on from the initial success of the mechanical NER cab signalling system, a more sophisticated electrical system was developed and tried out in 1907 (Fenner 2000, p3). The track equipment consisted of three triggering ramps, one centrally located and the other two located just outside of the running rails, 100 yards on the approach to a distant signal. These were followed by several centrally positioned ramps, one at the distant signal, several more ramps at intervals to the home signal, one at the home signal and further ramps at the following stop signals up to the most advanced starting signal. The locomotive was fitted with steel brush contact arms in the centre, wheeled contact arms at each side and a cab indication with a semaphore arm and route pointers (capable of pointing to 4 numbers).

As the train passed the triggering ramps the three contact arms were operated by the ramps. Operation of any of the three ramps caused a bell to sound in the cab and the semephore arm to indicate 'danger'. As the distant signal was passed, the central contact arm was again operated and, if the signal was clear, the cab indication changed to 'clear' and the bell stopped. However, if the distant was at caution the indication remained at 'danger' and the bell continued to operate until the driver stopped it by use of a cancelling leaver. At the following ramps the equipment was updated, allowing the indication to change if the aspects of the signal ahead had changed. The system was designed so that the cab indication acted as a distant signal. This meant that a clear aspect could only be given when the remaining signals in the station area were all displaying clear. On approach to a junction with a clear signal, the system would also provide a route indication (1 to 4) to indicate the route set (Kichenside et al. 1998, pp157-8).

An addition feature allowed the signaller to draw a train up to a starting signal at danger by sending an electrical signal to a train stopped over the home signal ramp. This caused the in cab semaphore display to flick to 'clear' and then back to 'danger', giving the Driver a 'call-on' indication (Kichenside et al. 1998, pp157-8).

Functionally, this system overcame many of the problems associated with its predecessors and later systems, but it was also very expensive. It was not, therefore, taken up beyond installation on trial sites and, by the 1920s, had been dismantled (Fenner 2000, p3)

C3.5 GREAT CENTRAL ATC (MILLER CAB SIGNALLING)

In 1903 the Great Central experimented with a system that provided stop and clear signal indications in the cab. This was based on transmission of signals to the train by induction from track circuits. It provided white and red lights, duplicated by a miniature semaphore, in the cab, but no supervision. At the time, the technology did not prove robust enough for the weather conditions prevalent at the trial site (between Crowden and Woodhead in the Pennines). Therefore, whilst the system was found to work fairly well, it did not progress beyond the trials (Dow 1965, p292; Fenner 2000, p4).

C3.6 THE STROWGER-HUDD SYSTEM

Following a series of investigations by the Ministry of Transport, a report was published in 1928 recommending the wide scale adoption of non-contact warning systems (Fenner 2000, p4). This resulted in the development of the Strowger-Hudd system, installed at distant signals, which replaced the ramp and plunger of the Great Western system by two magnets in the centre of the track and a receiver on the train. The first magnet was a permanent magnet, the second an electromagnet located 50 feet further along the track and only energised when the corresponding signal was showing a clear aspect. As a train passed the magnets, the flux of the permanent magnet was picked up by the trainborne receiver, which operated a relay to open the vacuum brake pipe to the atmosphere through a horn in the cab. If the signal was displaying a clear aspect the energised electro-magnet would then be passed, again operating the relay to close the vacuum pipe, releasing the brake and cancelling the warning. However, if the signal was displaying a caution aspect the horn continued to sound (and the brake remained applied) until the Driver operated a cancel handle (Kitchenside 1998, p161).

In addition to the audible warning and indication, the Strowger-Hudd system provided a visual indication to the driver by means of a circular display divided into segments. The display was all black, changing to alternate black and yellow segments after acknowledgement of a warning. This provided a reminder of the last aspect passed, which was a significant improvement on the Great Western system (Kitchenside 1998, p161). The use of electro magnetic signals also avoided the need for direct physical contact, hence reducing system wear and improving reliability (Hall(1) 2000, p159). However, there were also drawbacks - particularly from the fact that the same audible warning (a horn) was given for both clear and caution (Kitchenside 1998, p161). This limitation was extenuated by the reliance placed on correct driver operation to ensure safety (operation of the cancel handle also acting as an override to the brake application) (Hall(1) 2000, p159).

C3.7 BRITISH RAIL AUTOMATIC WARNING SYSTEM

Following nationalisation in 1948, British Railways decided to adopt a warning system for the whole network. A modified version of the Strowger-Hudd and Great Western systems, known as the Automatic Warning System (AWS), was subsequently installed on 98% of the UK main line network (Bailey 1995, p269). This system, unlike the earlier systems, was intended for installation at all signals. As with the Strowger-Hudd system, AWS requires two magnets mounted in the track on the approach to each fitted signal. The first magnet is a permanent magnet, with the south pole facing upwards. The second is an electromagnet with the north pole facing upwards and only energised when the corresponding signal is showing a clear aspect. As a train passes the AWS magnets, the flux of the permanent magnet is picked up by the trainborne receiver, which operates a relay. After a one second delay, this in turn cuts the current supplying an electro-pneumatic valve. Air is then allowed into the brake system and a horn is sounded. At this stage, the driver can acknowledge the warning and, by so doing return the system to normal. Failure to acknowledge the warning causes the brakes to apply (Barwell 1983, p106-7).

If the signal is displaying a clear aspect, the train passes the electromagnet, which in this system is located imediately after the permanent magnet, well within the allowed one-second delay even at 4mph (Loganathan 1993, p4). As the receiver detects the flux from this magnet, a bell rings for two seconds and the system returns to normal, thus preventing the sounding of the horn (Barwell 1983, p107).

As the AWS system provides an audible indication at every signal, it provides train drivers with location cues. Over and above the benefit that this has in alerting the driver to the state of the signalling, it can also be used in determining braking locations for station stops and speed restrictions ahead (Nichols et al. 1986, p10; also Appendix A interview with N Wright).

In bi-directional areas, a second 'suppresser' electro-magnet is fitted alongside the permanent magnet. This is energised for a move in the opposite direction, producing a magnetic field that is equal and opposite to the permanent magnet. In this way the AWS indication is suppressed (Loganathan 1993, p5).

The AWS system retains the segmented visual indication used by the Strowger-Hudd system, and therefore provides a reminder to the driver of the last aspect passed (Fenner 2000, p5). It is also a significant advantage of the AWS system that the track-based equipment will always give a caution indication in the event of a loss of power and that, similarly, "the AWS receiver is inherently fail safe. If the receivers north output is lost, e.g. through damage to the receiver, the caution sequence is automatically initiated" (Pincock 1998, pp237-9)

Whilst the AWS system offers improvements on both its 'Automatic Train Control' and Strowger-Hudd predecessors, it still has limitations. In particular:

- "The ability of the driver to override the automatic brake application could be said to be a serious dilution of its effectiveness as a safety device" (Rayers 1989, p13);
- It can only indicate two states – clear and caution. It is therefore unable to differentiate between Single Yellow, Double Yellow and Red aspects. This fact has some serious human factors implications, particularly when a train is running fairly closely behind a preceding train. In this situation the train may pass consecutive signals resulting in AWS warnings that require acknowledgement. This can lead to 'automatic behaviour' where the Driver acknowledges a warning without actually registering that it is for a more restrictive aspect than the preceding warnings. Alternatively, the driver may accidentally read the wrong signal and subsequently acknowledge an AWS warning in the belief that it is for a less restrictive aspect than it is. A number of accidents have occurred as a result of these types of situation (Muttram 2001, p1);
- In accordance with the rest of the system, the sunflower display also does not differentiate between Single Yellow, Double Yellow and Red aspects. AWS is therefore unable to help prevent drivers starting from a platform against a signal displaying a red aspect. This situation has also resulted in a number of accidents,

although AWS is now supplemented by the use of DRA, which has helped to reduce the number SPADs in these circumstances;

- In recent years, problems have been experienced with the reliability of AWS equipment, including automatic brake applications in inappropriate circumstances. These have been significant enough to undermine confidence in the AWS system and make isolation of AWS equipment a common practice (Fenner 2002, p31; Uff et al. 2001, pp29 & 63).

As a later development, following a derailment at Morpeth in 1967, the AWS system was also installed for use on the approach to permanent speed restrictions and, in 1975, for temporary restrictions.

C3.8 THE 'CROCODILE' SYSTEM

A system used extensively on SNCF, CFL, NMBS/SNCB and Eurostar, on lines operating between 40 and 160 km/h, BRS was introduced in 1930. It is based on an elecro-mechanical contact between ramps located in the centre of the track next to each trackside signal and a brush located on passing trains. The ramp (which resembles a crocodile, giving rise to the systems' more common name) is energised by a battery at –15V if the signal is green or +15V if it is at caution. The other polarity is connected to one of the running rails. Contact between the ramp and a trains brush therefore completes a circuit from the battery through the brush to on-board relays and back through the train's wheels and the running line to the battery. The on-board relays detect the polarity of the current flowing in the circuit and can therefore determine whether the signal ahead is clear or displaying a restrictive aspect.

A positively energised ramp triggers a flashing yellow light in the cab and the sounding of a short 900Hz tone. Following acknowledgement, the indication in the cab becomes steady as a reminder that the last signal was at caution. If the driver does not acknowledge by pressing a vigilance button (within 4 seconds on Eurostar, 5 seconds on SNCF and CFL or 7 seconds on NMBS / SNCB), an automatic emergency brake application occurs. This remains in effect until the train comes to a stand and the acknowledgement button is activated.

A negatively energised ramp causes a bell to sound briefly in the drivers cab. No visual indication is given and the driver is not required to take any action. However, if the previous signal was at caution, the steady indication lamp is turned off (Bailey 1995, p268 & 276-7; Pincock 1998, pp236-8).

The system is highly simple, yet sturdy and reliable, but its performance standards are limited – particularly with respect to maximum speeds (Guilloux 1991, p46).

Functionally, there is very little difference between the Crocodile system and AWS. The main differences are that:

- If the driver fails to operate the vigilance device before intervention occurs, the Crocodile system operates an emergency brake that can not be revoked;
- As there is no indication for a clear signal, failure of the trackside power supplies or of the train's brush causes no indication to be given (where AWS would default to a caution indication).

(Pincock 1998, pp238-9)

The similarity between the systems means that Crocodile has the same inherent weaknesses as AWS – It provides the same warning for all restrictive signals and can be overridden by the driver.

C4 AUTOMATIC TRAIN STOP SYSTEMS

In all of the systems described so far, the Driver is able to override the system to prevent a brake application. The systems were and are therefore unable to prevent a signal being passed at danger, or to mitigate the circumstances once a SPAD had occurred. An alternative approach to enforce SPAD protection was therefore developed in the form of Automatic Train Stop systems, which were defined in section 4.3.2, as 'systems automatically enforcing compliance with movement authorities'.

C4.1 MECHANICAL TRAINSTOPS

The most basic form of train protection to enforce SPAD protection is the mechanical trainstop. Mechanical trip arms were first trialled in the USA on the Fitchburg Railroad, Boston Elevated Railroad and Revere Beach & Lynn during the 1890s, with the first permanent installation being on the Boston Elevated in 1901 (Aldritch 1993, pp52-3)

In modern arrangements, 'trainstop' units are fitted adjacent to all stop signals and consist of a metal arm, which is raised when the signal is at danger so that the arm sits a few inches above rail level and just outside of the right hand running rail. When the signal is displaying a clear or caution aspect, the arm is lowered. If a train passes the signal when it is at danger, a pivoted arm mounted on the leading bogie collides with the trainstop arm and is pushed back. This cuts the current to the traction motors and opens an air valve to apply the brakes. After the train has been brought to rest in this way, the driver can reset the system by opening a door in the front of his/her cab and pulling a cord that restores the arm to its running position (Kichenside et al. 1998, p165). Lowering of the trainstop arm itself is typically powered by air controlled by an electro-mechanical valve and piston (on LUL this is obtained from the 60 PSI air main). Alternatively, an electric motor or hydraulic power pack can be used to drive a ram that holds the arm down. In all cases a heavy spring is used to return the arm to the upright position. This ensures that the default state (in case of power loss) is upright, making the device fail-safe (Loganathan 1993, p1).

When fitted to a bi-directional area, special arrangements must be made to ensure that a train is not improperly tripped by a trainstop intended to prevent movements in the opposite direction. "In this case the signalling design must ensure that the trip arm is lowered as the train passes" (Clark 1999, p2).

As this system is only able to stop a train after it has passed a signal at danger, trainstops are only fitted to stop signals, not to repeaters. Safety can then only be ensured by the provision of an overlap after the signal that is of sufficient length to guarantee stopping a train within it under worst-case conditions of speed and braking performance. The length of overlap required (typically 300m for a metro railway, but potentially more than 1500m for a main line railway) can therefore have a major effect on achievable headway. As a result, this system is not practicable for a high-speed railway (Fenner 2002, p31; Loganathan 1993, P1). The mechanical aspects of the system also limit the use of trainstops to low speed lines and, even at low speeds, they are prone to failure. Common causes of failure include burst hoses, leaking air valves, the internal mechanics becoming frozen by ice and the head becoming damaged. Over the years, means of protecting against wrong side failures have been developed, but mechanical trainstop failures nonetheless represent a significant problem to reliability.

As the mechanical trainstop system relies upon contact between a trainstop head and tripcock arm, both components must be carefully aligned (and regularly checked) in order to ensure correct operation.

To prevent trains passing signals at green but being tripped by a defective trainstop, the aspect controls are arranged such that, when a signal is ready to step up, the green aspect is illuminated first, whilst the red aspect is kept alight until the trainstop is proved to be lowered. The driver will see both aspects illuminated for the second or two it normally takes for a trainstop to lower. However, if both aspects remain lit, the driver understands that the trainstop has failed and will bring the train to a stand at the signal and consult the signaller. Conversely, a signal will not clear if the signal ahead of it is at red but its trainstop is not proved to be raised. This protects against trainstops failing in the lowered position (Clark 1999, p3; Dapre 1999, p5).

As the reset can be carried out by a driver alone, the trainstop system can not protect against a driver deciding to reset and continue after a trip has occurred. Whilst the railway rules should prevent this by requiring the driver to proceed ready to stop on sight of an obstruction until clear signals have been seen, history has shown that drivers do not always follow these rules. As a result there have been a number of serious accidents caused by a lack of caution in the application of 'trip and proceed' actions by drivers. In order to mitigate some of the risk presented by this situation, recent rolling stock refurbishment on LUL have seen the introduction of timers to the traction control system, enforcing a reduced speed of 8km/h to be maintained for two minutes after a trip reset (Dapre 1999, p5). This development has in turn created difficulties for the railway operators, as trains required to 'trip and proceed' during signalling system failures are often forced to continue at low speed after the driver has been able to ascertain from observed signal aspects that it is safe to resume normal working.

Mechanical trainstop systems are still widely used today on London Underground, as well as a few main lines branches in the UK.

C4.2 THE COTTRELL SYSTEM

Installed on the Liverpool Overhead Railway in 1914, the Cottrell system relied on contact between a pick up brush on the front of the train and a trackside contact just beyond the signal, which was energised if the signal was at danger. Detection of a voltage at a trainstop caused the train's traction motors to cut off and brakes to apply (Jarvis 1996, p76).

By depressing an override push button whilst passing a trainstop, the trainborne circuits were bypassed, allowing a failed signal to be passed. Misuse of this facility lead to two minor collisions in 1919 and a third in 1921. Accident investigations concluded that the override facility should have been located out of reach of the driver, such that it could only be operated by an inspector travelling with the driver (Jarvis 1996, pp77, 80).

The system was replaced in 1921 by a form of mechanical trainstop (Jarvis 1996, p80).

C4.3 THE GREAT CENTRAL RELIOSTOP SYSTEM

A more comprehensive system was developed in 1915 to provide a warning indication to the driver in conjunction with a trainstop system. This was done by means of a trigger point 300 yards before a distant signal. Short, side mounted ramps were located to the left of the track at both the trigger point and the distant signal. The trigger ramp was fixed and, when contacted by a vertical trip arm on the train, guided the arm in towards the track. This activated a pneumatic valve, admitting air to the vacuum brake system and sounding a siren in the cab to notify the Driver that he was approaching a distant signal. This warning was permanent (given regardless of the signals aspect). The Driver could then cancel the warning and brake application.

Additional equipment at the distant signal worked in the same way if the signal was at caution. However, when the signal was cleared the ramp was physically pulled clear of the contact position by the signal wire, and therefore gave no further indication in the cab.

At stop signals, the lineside equipment was located a few yards beyond the signal, so that it would not be engaged unless a SPAD occurred. The lineside unit consisted of a pivoted arm that was retracted when the signal was clear. When the signal was at danger, it pointed towards the track, at a height above the trigger point and distant ramps. A second arm on the locomotive would then engage with the trip arm and initiate a brake application that could not be revoked. Once the train had come to a stand, the Driver could reset the system by physically moving the trip arm back to its normal position and inserting a peg (Kichenside et al. 1998, pp160)

A total of 40 miles of line were fitted with the Reliostop system. However, the mechanical aspects of the system offered concerns for reliability, and further use was stopped when the Great Central was amalgamated with the LNER in 1923 (Dow 1965, p293; Hall[1] 2000, p93).

C4.4 SOUTHERN REGION AUTOMATIC WARNING SYSTEM (SRAWS)

In the late 1960s there was concern that trains were regularly operated on yellow aspects at most signals passed during peak times. It was recognised that this might lead to drivers making automated responses, thereby invalidating the protection that AWS was supposed to offer to them. As a result of this, an enhanced version of AWS was developed (Fenner 2000, p5). It provided two cab indication, one of the signal aspect that was being approached, the other of the aspect that was shown by the last signal passed (Alston et al. 1971, p29).

The system still used conventional AWS equipment, with the additional aspect information being transmitted to the train by use of an inductive wire loop. This was energised at a different frequency (in the range 410-620kHz) for each possible aspect. Loops were usually laid along the sleepers between each AWS magnet and the signal it was associated with, but could be extended to cover a longer distance. By this means the system could provide a longer sighting distance, or even continuous cab signalling, if required (Alston et al. 1971, p29-31)

When an AWS magnet was passed, the driver was required to acknowledge the aspect in the 'approaching' display (unless it was green) by operation of a button specific to that aspect. If the driver failed to acknowledge, or operated the wrong acknowledgement button, a cancellable brake application was initiated (as with failure to acknowledge a normal AWS warning). In addition to this, a trainstop function enforced an un-revokable brake application if the train passed a stop signal at danger (Kichenside et al. 1998, pp205).

This system was initially known as 'SRAWS' because it was tried out between Southampton and Bournemouth on the southern region, but subsequently became known as 'signal repeating AWS'(Kichenside et al. 1998, pp205-6).

Despite the advances in protection offered by SRAWS, the system did not progress beyond the trials. This was largely due to the expense of the reed transmission technology that it used (Fenner 2000, p5).

C4.5 THE 'SIGNUM' SYSTEM

Another system similar to AWS and Crocodile is the 'Signum' system used in Switzerland. Introduced in 1934, Signum can be used at speeds of up to 160km/h. Like AWS it is based on induction between magnets. Two coils in the track are connected together if the signal is at danger and the connection is broken if the route is clear. Two further coils are mounted on the train, one being permanently magnetised (Bailey 1995, pp276-7).

When a signal is at danger, the permanently energised magnet on the train induces a flux in one of the track mounted coils. The resulting current flows to the second track mounted coil, causing it to induce flux in the second train mounted coil. The induced current on the train is then used to operate controls (Barwell 1983, p108).

Since developments in 1976, Signum is capable of providing separate 'stop' and 'warning' aspects, overcoming some of the major limitations of AWS. The system is further enhanced by trainstop functionality, with the emergency brake being automatically applied when the driver passes a stop aspect. Since a normal approach will have given the driver a warning at the distant signal, the trainstop function is activated without warning (Bailey 1995, p269).

However, Signum provides no transmission for a clear signal and can not detect failure of a coil, which means that it is not fail safe.

C4.6 TRANSMISSION BALISE LOCOMOTIVE 1 (TBL 1)

First introduced in 1987, TBL is installed on part of the NMBS/SNCB network - including some lines that are also fitted with Crocodile. Where trains are fitted to receive signals from both systems, TBL is given priority. If, however, a TBL signal is not received, trains respond to Crocodile signals.

This system operates at speeds of up to 160km/h, using beacons that transmit signals to passing trains based on the current aspect of associated signals. The beacon is located off centre in the track, to allow automatic detection of running direction. It continuously emits a message in the 100kHz range, consisting of five data codes. Each code is selected from ten available codes (giving a total of 100,000 possible code combinations). When the associated signal is displaying a stop aspect, the beacon transmission automatically triggers an emergency brake application. When associated with a warning aspect, the transmission triggers an audible warning. If the driver does not acknowledge this, the brakes are automatically applied (in a similar sequence to that of Crocodile).

In addition to the trainstop and warning functionality, a display in the cab indicates whether the last signal passed was a stop signal or shunting signal.

TBL 1 protects against system failure by utilising two out of three computers and automatically triggering application of the trains brakes if a telegram is not received (Bailey 1995, p270-1 & 276-7; Pincock 1998, p237 & 243).

C5 PARTIAL ATP SYSTEMS

The main limitation of all Automatic Train Stop systems is the fact that intervention can not be enforced until a signal has actually been passed at danger. In addition to this, they provide no monitoring of train speed and can not, therefore, intervene if the speed becomes excessive. These limitations have been overcome in part by the development of systems also capable of automatically enforcing compliance with speed restrictions and movement authorities at some locations, such as on the approach to a signal. This type of system was defined in section 4.3.2 as 'Partial ATP'.

C5.1 THE 'INDUSI' SYTEM

Induktive Zugsicherung (Indusi) was developed by Siemens in the late 1920's and has seen widespread use by DB in Germany, ÖBB in Austria and the Yugoslav Railways, on lines with a maximum speed of 160km/h. Indusi, utilises three resonator coils or 'beacons' in the track. One is located at the warning signal (typically 1000m before the stop signal), tuned to 1000Hz. The second is tuned to 500Hz and typically located 150 to 250m before the stop signal. The final beacon is tuned to 2000Hz and located at the stop signal (Barwell 1983, p107-8; Bailey 1995, p269 & 276-7; Lange 1975, pp203-4).

A frequency generator on the locomotive continuously generates the same three frequencies in open resonant circuits (Geduhn 1995, p47). When the signal is clear, the beacon coils are short-circuited to prevent resonance. A passing train therefore receives no indication (Bailey 1995, p269). When the signal is at danger, however, the beacon resonant circuit draws energy from the appropriate resonant circuit on the passing train. This causes a relay on the train to drop out and register the passing of a beacon at that frequency (Geduhn 1995, p47).

After passing the warning signal, the driver receives an audible and visual warning, in response to which he/she must operate a vigilance control. Failure to do so within 4 seconds causes a full brake application to occur. In addition to this, the INDUSI system enforces a reduction in speed following the warning signal. An on-board switch allows selection of a time delay of 20, 26 or 34 seconds, by which time the train must have reduced its speed to 95, 75 or 60km/h respectively. This can be used to adapt supervision to the class of train (fast, slow or freight vehicles). Failure to comply with the required speed reduction also causes a full brake application. Similarly, the train must be travelling at a speed below a pre-defined ceiling (40, 50 or 65 km/h, depending on train type) before the second beacon is passed, and must not pass the beacon at the signal itself. Failure to comply with any of these requirements again causes a full brake application. (Barwell 1983, p107-8; Bailey 1995, p269 & 276-7; Lange 1975, p203).

An override facility is provided to allow a signal to be passed at danger. When operated, this enforces a speed limit of 40km/h on all train movements. In the event of a trainborne equipment failure, operating rules require the driver to keep train speeds below 100km/h on DB AG and below 120km/h on ÖBB (Bailey 1995, p269 & 277).

Indusi is also used to enforce speed restrictions. A permanently active 1000Hz coil is located on the approach to the speed restriction, and a 2000Hz coil at its actual commencement. The1000Hz coil initiates a warning, as with approach to a stop aspect, and also starts a timer connected to the 2000Hz coil. On expiry, the timer deactivates the 2000Hz coil. If the train arrives before this deactivation, it is considered to be travelling too fast for the restriction and its brakes are automatically applied (Lange 1975, p204).

In the INDUSI system, no power is required by the track-mounted coils. The system also has advantages over AWS in that it provides different functionality for the stop and distant signals and provides speed supervision at the start of a speed restriction, rather than just a warning. It has the disadvantage, however, that it is not fail safe. If one of the track based coils is removed or damaged, no warning is given and the system behaves as though the signal ahead is clear. (Barwell 1983, p107-8; Lange 1975, 203).

An advanced version of INDUSI (known as INDUSI 60 Rechner – I60R) was introduced in 1992. This system provides semi-continuous speed supervision on the approach to a signal at danger (Bailey 1995, p270).

C5.2 ASFA

An equivalent system to INDUSI used on RENFE lines is known as ASFA. This system was introduced in 1972 and uses beacons to transmit 5 frequency codes between 60kHz and 111kHz. The codes are used to represent signal aspect and target speed information. ASFA is used at speeds of up to 220 km/h and has the same basic functionality as INDUSI. On passing a beacon on approach to a stop aspect, the beacon transmission is used to perform a speed check. If an overspeed is detected, the emergency brake is automatically applied. When the transmission indicates a restrictive aspect, a warning is given to the driver (in the form of a bell). Following this warning signal ASFA gives the driver 3 seconds to operate the acknowledgement button before automatic application of the emergency brake.

Unlike INDUSI, this system has a separate transmission for clear signals. Despite this, however, the system is not fail safe and can be switched off without any restriction to train running (Bailey 1995, p270 & 278-9).

C5.3 TRAIN PROTECTION AND WARNING SYSTEM (TPWS)

In 1994, following the acceptance by the HSE and the Secretary of state that it was not reasonably practicable to fit either of the BR ATP systems nation-wide [see section C6.1], it became necessary to consider alternative schemes. As a result, a working group known as the Signals Passed at Danger Reduction And Mitigation (SPADRAM) group made two technical proposals. One was for a Drivers Reminder Appliance (DRA – see section C2) and the other for an enhanced form of AWS. In response to these proposals, an enhanced AWS specification was developed. The key aims of this specification were to keep costs down by minimising the changes that would be required to trains and by ensuring that the system could be implemented at selected high-risk locations without needing to be installed at other locations (Fenner 2000, p7). This specification resulted in the development of the UK's main example of a partial ATP system, TPWS.

TPWS utilises loops, each about 1m in length, that are installed in four foot of the track in pairs. These loops are driven by tone generators that are activated by signalling relay contacts when the signal is red. The first loop in each pair 'arms' the system and the second 'triggers' it (Fenner[(1)] 2001, p2-3). The tones that they transmit are detected by an antenna on the train, which is installed as part of a logic unit that replaces the old AWS unit (thus simplifying train fitment) (Fenner 2000, p8).

Loops are located abutting each other at critical signals, in order to provide a trainstop function. A train passing these loops whilst the signal is displaying a red aspect detects the 'arming' tone immediately followed by the 'trigger' tone. This causes an automatic application of the train's brakes. If, however, the 'trigger' tone is detected followed by the two tones occurring simultaneously, the system recognises a reverse direction move and ignores the trainstop signal (Fenner[(1)] 2001, p2-3).

Additional loops are located on the approach to the signal in order to provide overspeed supervision. These are separated by a distance that would take a train one second to travel at the speed that they are to enforce. The first loop starts a timer on the train which must expire before the second loop is detected. If it has not done so, the train's brakes are applied. The overspeed timers installed on freight trains run for longer, effectively reducing the value of the supervised speed by 25% in order to account for their different braking performance. Once applied, the brake intervention can not be released for 60 seconds (Fenner[1] 2001, p2-3; Stretton 1999, p417).

In total, 6 frequencies in the 60 to 70 kHz band are used to provide TPWS functions. Three are used for overspeed and trainstop loops in one direction (for arming, trainstop trigger and overspeed trigger) and the other three for the other direction. This allows the system to tell the difference between a trainstop and overspeed trigger signal and also permits full operation in bi-directional areas (Fenner[1] 2001, p2-3).

TPWS overspeed loops can also be fitted to other speed critical locations, such as the approach to a speed restriction or buffer stops (Stretton 1999, p417). However, when used in this way "its effectiveness is likely to be limited because the difference between an acceptable speed of approach and an unsafe speed is small" (Fenner[2] 2001, p4).

Additional driver controls are provided to:

- Permit a train stop (but not an overspeed sensor) to be overridden when the driver is instructed to pass a signal at danger;
- Enable isolation of the system in the event of failure or for special operational needs;
- Indicate the operational status of TPWS.

(Fenner[3] 2001, p2)

TPWS offers considerable advantages over an AWS / DRA type of system. In particular, it offers a considerable reduction in the risk of collision or derailment due to a SPAD or overspeed and, unlike AWS, the driver is unable to override an intervention. However, it also has serious limitations:

- Even if installed at all signals, TPWS only offers intermittent protection at the discrete locations where trackside equipment is installed. In practice, it is only planned to install the system in critical locations (about 40% of all signals), which will leave most of the rail network unprotected (Stretton 1999, p417).
- When installed, the location of speed measurement loops gives the system a fixed, one off, speed measurement. This is typically just over 300m before the signal, with the intention of stopping trains travelling above 66mph. Those travelling between 66 and 75mph will be stopped before the end of the overlap, but trains travelling at higher speeds will not. Similarly, the effective speed at which a trainstop can stop a train before the end of a standard 183m overlap is only 40mph (Fenner 2000, p11). This means that there are 'gaps' in the speed supervision coverage of TPWS at speeds above 75mph or between 40 and 66mph. These gaps are made worse by the fact that trains can stop following the required braking curve, or even accelerate, after passing the overspeed loops (Fenner[2] 2001, p4).

The main line railways for which TPWS is designed also utilise a variety of rolling stock types, with significantly varying braking performance. As a result, there is not a single optimal loop location for all trains, and the optimal location for one stock type could result in trains of another type being unnecessarily tripped during their normal braking curve. The choice of location is therefore a trade off between safety and efficient operation of the railway, taking into account the performance of all stock that may use the line.

The reliance placed by TPWS on energised loops to send signals to passing trains means that it is not a fail safe system (Dapre 1999, p8). An argument (with some considerable validity) in defence of TPWS notes that it is a supplementary safety system that does not take on the primary responsibility for safety, which remains with the driver. It is transparent to the driver, which means that it can not lead him/her to make an incorrect action and, in consequence of this, does not need to be fail safe (Muttram 2001, p4). However, in order to maximise the safety that can be afforded, it has been designed to include some monitoring and fault alarms:

- TPWS trackside equipment monitors the current flowing in the loops and sends failure alarms if they are not energised whilst the signal is at red. As a part of this alarm function, the signal in rear is returned to danger so that approaching drivers can be stopped and warned about the lack of train protection at the signal ahead. (Fenner[1] 2001, p5).

- TPWS trainborne equipment goes through a diagnostic self test every time the driving cab is powered up (the results of which are confirmed to the driver). (Fenner 2000, p10)

The provision of this monitoring capability was not originally intended in the enhanced AWS specification (which was supposed to result in a simple system). Unfortunately, TPWS developed "a number of teething problems during trials which, whilst progressively overcome, have led to greater complications of the equipment" (Uff et al. 2001, p30), the monitoring capability being an example of this. The addition of these facilities has caused the cost per signal for TPWS to become "similar to the cost of *{BR}*ATP" (Fenner[2] 2001, p1). Whilst TPWS does not need to be fitted to all signals, meaning that it is still significantly cheaper than BR ATP (see section C6.1), the gradual increase in TPWS cost and complexity has lead to criticism of the design.

The TPWS system is designed to avoid, or at least mitigate, the most serious SPAD incidents, where a train overshoots the signal at speed and causes a collision or derailment hazard. "In general it will not reduce the number of SPADs. A SPAD will already have occurred before the trainstop activates. Even with an overspeed sensor the speed settings must be high enough not to intervene when the driver is driving correctly. A consequence is that, if a speed trap does intervene, a train even if just above the threshold is likely to pass the signal by a small margin" (Fenner 2000, p8). Initial predictions suggested that TPWS would prevent about 70% of ATP preventable accidents, with its effectiveness being least where the risk is greatest - on high-speed trains with below optimum braking capability (Dasi-Sutton 2000, p23; Uff et al. 2001, p107). These predictions were based on the assumption that any collision resulting from SPAD incidents would occur at the end of the signal overlap. However, subsequent analysis has shown that many SPADs would not result in a collision at the end of the overlap, but rather at a distance beyond it. This reduces the predicted speed of any collision that would occur and also makes it more likely that the train would stop before the collision point. The predicted effectiveness has, therefore, been increased to 81% of ATP preventable accidents (Waboso 2002, pp7, 9 & 40).

Doubts have also been raised as to the reliability of the TPWS system. By 1999, testing showed that TPWS "had not yet achieved reliability comparable to that of AWS", a system that is itself noted for an "underlying level of unreliability" (Uff et al. 2001, pp 60-3). The installation of TPWS also had negative effects on the reliability of existing BR-ATP equipment on-board the fitted trains (Wright 2004, p9)

Inconsistencies have also developed in the ways that TPWS responds to failures. "Contrast the growth in cost and complexity of what was supposed to be a simple system, to ensure that a failed loop is protected by the signal in rear, with the lack of similar protection should the trainborne TPWS equipment fail in service. There is no TPWS status indication in the cab nor a warning that the equipment on the train has failed" (Ford[1] 2001, p12). Clearly, once the driver is aware of a trainborne failure, operating rules will require the train to be removed from service, but whilst the driver is unaware, operation will continue without protection (Uff et al. 2001, p63-45).

Conversely, a key advantage of TPWS is that it can be targeted, and rapidly deployed, in priority areas (Sullivan 1999, p907; IRSE[3] 2001 section 9). Whilst it will only marginally reduce SPAD events in these areas, "it will limit the severity of many more since trains will often stop in the overlap" (Fenner[2] 2001, p1).

In some areas, TPWS has been installed on lines already fitted with BR ATP (see section C6.1). Although TPWS is a lower integrity system than BR ATP, it has been found to sometimes be more restrictive. "This can cause conflicts of information to be given to the driver, or degradation in performance (e.g. an involuntary stop)" (Tomlinson 2001, p45). As a consequence of such performance problems experienced with TPWS, particularly on approach to buffer stops, driver confidence in the system was erroded (Wright 2004, pp10 & 12).

Whilst implementation of TPWS is now underway across the UK rail network, the Southall and Ladbroke Grove joint enquiry into train protection systems concluded that: "after weighing the issues carefully, we are left with considerable reservations about the effectiveness of TPWS... Its limitations have become more apparent as its cost has escalated. We doubt that it would have been adopted had both of these factors been known more reliably at the outset" (Uff et al. 2001, p107).

TPWS was due for fitment to all trains and 46% of signals on the UK Main Line network by the end of 2003, and has a design life of 15 years (Waboso 2002, pp7 & 9). The cost per life saved associated with implementation of the system across the UK network is quoted as in the region of £3 million (Hall[2] 2000, pB6-9).

C5.3.1. TPWS +

TPWS + is the term that has become applied to the extension of the operating range of TPWS to make it effective at higher speeds and/or lower braking capacity, by placing extra overspeed loops further away from the signal. The provision of loops set for 75mph, approximately 770m before a stop signal with a 183m overlap, would provide protection at up to 100mph with 12%g braking (Davies 2000, p64). Further protection could then be provided by provision of a third or more pairs of loops (Uff et al. 2001, pp70-1).

In accordance with the proposal to implement TPWS +, the manufacturers of TPWS have confirmed that "the equipment will operate at up to 140mph and that the speed setting at a given location can be adjusted simply by spacing the loops further apart" (Sullivan 1999, p907).

TPWS+ with one additional pair of overspeed loops was initially expected to be about 75% effective on preventing ATP preventable accidents – an improvement of 5% (Dasi-Sutton 2000, p23). Based on this prediction, Sir D Davies supported the adoption of TPWS+ in his 2000 report for the UK government on train protection systems (Davies 2000, p65). However, by the following year, the Southall and Ladbroke Grove joint enquiry expressed reservations about this approach, noting that whilst TPWS+ could be used to improve safety in critical location, this would further increase the cost of TPWS installation. "We believe that it is important to keep well in view the original objective of TPWS, which was to provide a quick and cheap stop-gap solution...The risk that the partially effective system (TPWS) may delay implementation of the fully effective system (ETCS) can only be exacerbated by deployment of enhanced TPWS systems" (Uff et al. 2001, p75). The same report also raised concerns that the provision of any additional loops may result in operational problems for trains with different braking characteristics (Uff et al. 2001, p110).

More recent analysis by the ERTMS Programme Team has lead to an adjustment in the prediction for TPWS+ effectiveness to 83% of ATP preventable accidents - an improvement of only 2% on their revised prediction for TPWS. This small improvement over the basic system arises due to the fact that very few accidents result from SPADs at full train speed (Waboso 2002, p41). In the light of these revised predictions, TPWS+ appears even less attractive than it did at the time of the Southall and Ladbroke Grove joint enquiry.

TPWS+ is currently undergoing trials and the extent (if any) of its future implementation on the UK rail network is not known.

C5.3.2. TPWS-E

A further development of TPWS, this system is proposed to replace the track loops with Eurobalise beacons (see section C6.7). The trainborne equipment could then be identical to that required for ETCS systems.

TPWS-E would be expected to offer a number of advantages over TPWS, including:

- Allowing the possibility of subsequent migration to ETCS systems;
- Open procurement and multi-sourcing;
- Simplified interfaces to existing signalling (the ETCS technology uses current sensors to determine signal aspects, rather than connection to spare relay contacts);

- Inherent detection of failed balises (ETCS balise transmissions can contain linking information to advise of the location of balises ahead);
- The greater transmission capacity of Eurobalise technology means that operational advantages could be gained by simultaneous transmission of several messages for different types of trains;
- Reduced cabling requirements.

  (Davies 2000, p65; Uff et al. 2001, p74)

TPWSE is again currently undergoing trials and the extent (if any) of its future implementation on the UK rail network is not known. However, it would, seem likely that TPWS-E will be adopted in preference to TPWS if the trials prove successful.

C6 INTERMITTENT ATP SYSTEMS

As explained in section 4.3.2, Intermittent ATP is a sub-category of Comprehensive ATP. It is therefore 'a system automatically enforcing compliance with **all** speed restrictions **and** movement authorities (for all vehicles) within a given area'.

With conventional lineside signalling, the signals act as an intermittent source of information on which the driver bases his/her actions. Having passed one signal, drivers receive no further updates until the next signal is seen. Intermittent ATP systems act in a similar way, with the trainborne equipment providing continuous supervision against overrunning a signal and / or exceeding known speed restrictions, but with the information that it supervises against only being updated intermittently from the trackside. All intermittent systems suffer from being over restrictive in normal operation, since they are unable to respond to a less restrictive movement authority (the equivalent of a driver observing a signal aspect clear) until the next transmission point is passed. As a result, the capacity of the system depends heavily on the number and designed location of transmission points.

An example of the impact of different localised transmission arrangements on an Intermittent ATP system is shown in Figure C 6-1.

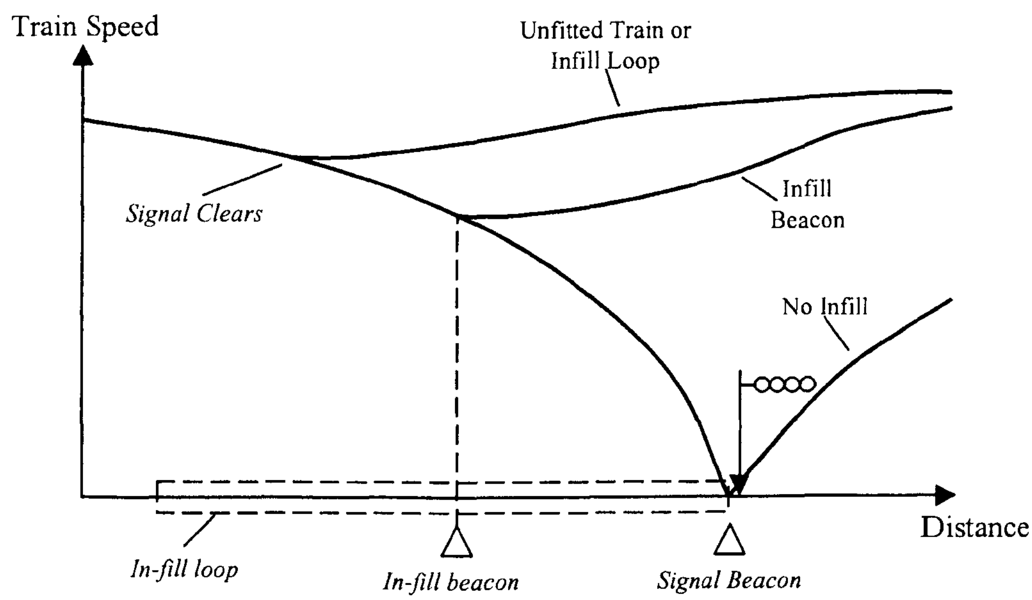


Figure C 6-1: Trajectory of Train Approaching a Red Signal with Intermittent ATP (based on Leach M 1991, pp266-7)

It is worth noting that the UNISIG ERTMS/ETCS specifications refer to this type of system as 'Spot Transmission'. However, this term does not seem to have attracted use or understanding as widespread as 'Intermittent ATP' and the author has therefore decided to continue with the use of the term 'intermittent'.

C6.1 BRITISH RAIL ATP

In the late 1980s, British Rail (BR) decided to undertake a significant trial programme for ATP on a part of the operational railway (Davies 2000, p17). A specification was therefore produced to define the operational and technical performance expected and to outline the constraints within which the system would operate. The main aims of the specification are to:

- Supplement, not replace, existing fixed block signalling;
- Be suitable for fitment to all BR lines and the whole BR traction fleet;
- Be capable of fitment either comprehensively or selectively at locations of high risk;
- Provide protection without detriment to normal train running performance;
- Be compatible with existing operational and technical practices, rules and procedures;
- Keep down costs by using intermittent track to train communication, mainly in the vicinity of signals;
- Calculate speed distance profiles onboard individual trains, rather than as a generic calculation trackside, in order to allow optimum performance of trains with different performance levels.

(Bailey 1995, p275).

The specification also requires the ATP system to be based on available off-the shelf technology and to provide:

- Speed supervision against the maximum speed permitted for the current train location. An audible warble and visual warning to be given to the driver if this is exceeded by 3mph or more and an automatic full service brake application if it is exceeded by 6mph or more;
- Braking supervision to ensure that the train will slow down sufficiently to achieve a target some distance ahead. To be achieved by calculating a target speed curve and then treating this in the same manner as a maximum permitted speed (an audible warble and visual warning being given to the driver if it is exceeded by 3mph or more and an automatic full service brake application if it is exceeded by 6mph or more);
- An indication to the driver of the current or target speed limit. Any new target being displayed in time to give the driver the equivalent of 3 seconds running time at the current line speed before a warning for under braking would be required at the current actual speed;
- A train trip function to provide an automatic emergency brake application that cannot be released until the train has come to a stand if the train is detected to have passed a signal at red.
- A train trip override facility (achieved by operation of a button whilst the train is stationary) to enable passing of a signal at danger under authorised conditions. The train trip function to then be re-enabled when the trainborne equipment detects that the signal has been passed or after the train has travelled 100m;
- Rollaway protection to apply an emergency brake application if the train is detected to be moving in the opposite direction to the driver's control setting.

(Holgate et al. 1993, p1-2, & 4-7; Appendix A Interview with Tim Brockbank).

In response to this specification, two trial projects were funded. One (known as the SELCAB system) was installed on the Chiltern Lines. The other system (based on TBL) was installed on the Great Western Main Line. Both systems are intermittent, using loops or beacons to communicate to the train information on position, gradients, speed limits, the signal ahead and the next beacon / loop to expect (Kichenside 1998, pp206-7). The contracts for these projects were let in 1990, with installation commencing in 1991 (Davies 2000, p17).

In both systems, there are three main operating modes:

- When track data is received for a signalled movement the trainborne equipment verifies that it is for the direction that the train is moving in and then determines the position of the train in relation to that track data. When this has been done, the trainborne systems operate in 'full supervision' mode (providing speed and distance supervision);

- Whenever verified track data is not available, but the trainborne system does know all necessary train specific data, the trainborne systems operate in 'partial supervision' mode (only providing speed supervision against the maximum permitted vehicle speed);
- When the required train data is not available, the train systems operate in 'shunting' mode (providing speed supervision restricting movement to 20mph).

  (Holgate et al. 1993, p5-6).

The same cab display was used for both projects, based on a conventional analogue speedometer fitted with LEDs around the dial and a three character dot-matrix display. In full supervision mode a green LED is lit to indicate the current speed limit, a flashing green LED to indicate a target speed limit and a steady yellow LED to indicate a release speed. In addition to this, the dot-matrix display gives an indication of the supervised signal stopping point:

- '===' for not being supervised to stop at a signal at danger;
- '..0' for supervised to stop at the next signal but two;
- '.00' for supervised to stop at the next signal but one;
- '000' for supervised to stop at the next signal.

During a brake intervention or warning a warbling audible alarm sounds, the LED is extinguished and the dot-matrix display flashes the current supervised speed, target speed or supervised signal stopping point (as appropriate for the cause of the brake intervention). Whenever a change occurs in the display, a 'blip' is sounded in the cab. In other modes the LEDs are not displayed and relevant symbols or mnemonics are shown on the dot-matrix display (Holgate et al. 1993, pp7, 9 & 11).

Because both systems are based on intermittent communication, the ATP equipment onboard the train can not always respond immediately to changes in the trackside signalling conditions. Where a signal is replaced to danger in an emergency, this means that the ATP system may be unable to provide protection for the train. As a result, the driver must remain vigilant for changes in the signalling conditions and can not rely entirely on the ATP system. In the case of a signal clearing to a less restrictive aspect, it does not affect the safety of the train movement, but does impose restrictions on its operation. The effect of this is most severe on approach to a signal displaying a red aspect, where a train could be forced nearly to a stand before an update is received, even though the signal cleared well before arrival. In order to overcome the worst part of this scenario, when the train reaches an appropriate position, the supervisory constraint is required to change from braking supervision to supervision of the train to a release speed. This speed is calculated onboard the train such that, if the train passes the signal at red, the trainstop function would still be able to bring the train to a stand within the available overlap. During supervision to a release speed, the value of the speed is displayed to the driver (Holgate et al. 1993, pp6-7).

The implementation of both schemes encountered difficulties, particularly related to retrofitting existing installations (both trains and track) and the management of temporary speed restrictions. "In the case of the Great Western route this led to extremely low reliability resulting in the system not being brought in to regular service until 1998. Even the new trains and signalling of the Chiltern lines were not immune to reliability problems, albeit far less severe" (Fenner 2000, p6). Both schemes also suffered from speed and distance discrepancies in periods of poor adhesion. This was particularly severe on the GWML fleet, where all axles are motored. On the Chiltern line, where the tachometers could be mounted on un-powered axles, the problem was less severe (Wright et al. 2002, p6).

As the first comprehensive ATP system to be widely fitted to main lines in Britain, the introduction of BR-ATP was unpopular with many drivers who saw it as a removal of their responsibility. They also complained that the audible alarms distracted them from their driving tasks (Wright 2004, p10).

By 1994, both ATP trials were deemed to have become disproportionately expensive for the benefit gained. As a result of this, the HSE and Secretary of State accepted that it was not reasonably practicable to expand the fitment programmes (Fenner 2000, p7). This view was partially supported by the 2001 Southall and Ladbroke Grove joint enquiry into train protection systems:

- The enquiry recommended that the schemes should not be extended to cover other lines, the ETCS systems be adopted instead for future projects;
- In accordance with the specified requirement for the BR ATP systems to be capable of fitment either comprehensively or at selected locations, short sections of track were omitted from the Great Western trial project. It was recommended that these should now be fitted to provide coverage over the whole line.
- It was further recommended that unfitted trains using the BR-ATP scheme lines should not be retrofitted with ATP equipment.

(Uff et al. 2001, pp104-5)

C6.1.1. GWML ATP (TBL)

The GWML ACEC system is based on the Belgian TBL system, using beacons and infill loops (Fenner 2000, p6). The beacons consist of a single turn of stainless steel 30cm wide and typically 1.5m long (or longer if required due to message length or operation at high speeds). They are located at linesided signals, off centre in the track. Only one beacon is required per signal but, where necessary, an infill beacon is also located a few hundred metres before the signal to allow for anticipation of an aspect change. As an alternative arrangement, an infill loop can be located on the approach to a signal for more continuous update to approaching trains. Where used, the loops include regular crossovers of the cable (about every 50m) in order to reduce crosstalk. Each transmission device emits a continuous digital message by Frequency Shift Keying at 90 and 100kHz (Binard et al. 1992, pp2, 4 & 5-6).

One antenna is installed at the leading end of the train (or one at each end for bi-directional working). The antenna unit contains two receiver coils, located such that one will pass directly over each of the beacon's longitudinal bars. By this arrangement the phases of the two received signals can be compared and rejected if a phase shift is detected corresponding to a beacon being for the opposite direction (or even another track). As the phase shift rejection method used for beacons cannot be used for loops, each loop is assigned an identification number that is transmitted to the train within its signal and also by the preceding beacon (Binard et al. 1992, pp4-5).

Beacons can transmit data at 25kBits per second and loops at 1kBit per second. The transmitted messages are formed of 8 blocks, totalling 230 bits. 97 of these bits provide usable information whilst the rest provide redundancy and synchronisation. A typical message contains:

- An identification number and type of beacon or loop;
- Linking date (distances and / or identifications) to the next balise or loop;
- The aspect of the related signal;
- The distance to the next target and the type of target (this is usually a signal). This includes whether the target is absolute or permissive, overlap lengths, etc;
- The authorised speed profile ahead, including the type of restriction (e.g. line speed, permanent or temporary speed restriction) and the distance to its start and finish points. Where different restrictions apply to different types of train they are all transmitted and the trainborne systems determine which applies;
- Track gradients;
- Other miscellaneous information (such as level crossings, ATP area exit, radio switching, etc.).

(Binard et al. 1992, p5)

A wayside encoder unit selects the transmitted data. This interfaces with the existing signalling by use of current transformers measuring the signal lamp currents. It then selects a message for transmission, based on the detected aspect (Binard et al. 1992, p6; O'Neill 1999, pp41-2). The encoder unit contains duplicated processors, both of which must generate identical telegrams before any message can be transmitted (Wright et al. 2002, p2).

The trains are fitted with a 2 out of 3 computer system that processes the received data and generates a speed profile for the train. Each channel carries out continuous self-tests, which are supplemented by a more rigorous test whenever the vehicle is brought into service and during maintenance (Binard et al. 1992, p7; Fenner 2000, p6). Fixed train parameters are stored in data plugs within the trainborne equipment, whilst variable train parameters (such as the train's journey, braking capability and length) are entered through the driver's data entry unit. One exception to this is the wheel diameter, which is entered through coding wheels in the trainborne equipment cubicle following re-profiling (Binard et al. 1992, p8).

The GWML scheme can notify the driver of emergency speed restrictions by use of a programmable parameter plug, containing a PROM. This is normally stored in the location case and can be quickly fitted to the encoder unit in the case of an ESR being imposed. However, it simply causes the letters 'ESR' to be presented on the drivers display, whilst supervision continues to the normal line constraints (Wright et al. 2002, pp2 & 5).

Where the signal separation is near to actual braking distance, processing delays result in the ATP system needing to know about signal aspects before reaching the first restrictive aspect. On the GWML scheme, which was primarily based on beacons, this caused unexpected problems with a significant number of additional infill beacons being required. The same problem was encountered on the Chiltern Lines, but easily overcome by extending the existing loops (Holgate et al. 1993, p7).

Another problem encountered on the GWML scheme was the reliability of the tacho generators. The 165 fleet and HSTs of GWML were fitted with axle end mounted tacho generators that relied on mechanical rotation to detect the trains movement. These included bearings which became worn by the forces that apply to an axle as a train moves along the track (degrading rapidly once the bearings started to wear). Initial errors in the installation design also led to water ingress, excentuating the wear problems. In consequence of this, they were very prone to failure (See Appendix A, interview with Tim Brockbank and Wright 2004, p9).

Where used, infill loops were found to sustain frequent damage and require lengthy repair times. Whilst not directly causing safety problems, this lead to reduced performance and driver frustration (Wright 2004, p12).

Further reliability difficulties were experienced due to unexplained self test failures on start up. Investigation suggested that this was due to the presence of interference during a closed loop assessment of the transmission path (when a very small signal was injected into the ATP antenna). In order to minimise the disruption caused by such failures (which occurred in up to 5 of the 9500 self-tests every month), the self-test has now been reduced from its initial 5 minute duration to 90 seconds, which allows time to re-run the test if it does fail . A re-run is now automatically initiated if the test fails and a failure is only reported if the r-run also fails (Wright 2004, p8).

Whilst specified and designed to operate as an intermittent ATP system, in practice not all trains running over the GWML fitted areas are equipped to operate within the protection of the ATP systems. As a result of this, the GWML BR ATP scheme could actually be viewed as Partial ATP (see definitions in section 4.3.2). Whilst fitted and unfitted trains continue to run interleaved within the ATP fitted areas, the safety of fitted trains and their passengers can not be guaranteed (since an unfitted train could be involved in a SPAD and crash into one of the fitted trains – as happened at Ladbroke Grove on October 5th 1999). Unfortunately, this situation is being maintained by the planned introduction of new rolling stock that will not be fitted with ATP equipment (Webster 2001).

C6.1.2. CHILTERN ATP (SELCAB)

SELCAB is based on the LZB system, with transmission loops laid out in front of each signal. The loop lengths vary between 5 (for shunt signals) and 300m, depending on the operational requirements at that location. The length is selected in order to provide for continuous reception of information updates on approach to, and when standing in front of, a signal at danger (Uebel 1996, pC1/12).

The system is actually capable of implementation in three levels:

- Level 1 uses intermittent transmission of data to trains, from local equipment associated with each signal, as the trains reach the signal. Level 1 has 5, 25, 50 or 100m loops;
- Level 2 extends the range of transmission to trains to cover the normal sighting distance of the signal (thus providing immediate release of braking supervision if the signal clears as the train approaches). Level 2 has 300m loops;
- Level 3 employs continuous data transmission to and from trains via a succession of 300m loops, with crossovers every 100m.

  (Barnard 1991, p294, 296).

The implementation for BR uses a combination of levels 1 and 2 (for which the trainborne equipment is identical), whilst level 3 would require additional hardware in the on-board equipment and corresponds to the LZB system (see section C7.4).

The lineside signal aspects were initially picked up by a high resistance connection to the signal head feed (Barnard et al. 1992, p7). A later Chiltern Line upgrade scheme adopted a different interface mechanism, obtaining a direct output from an SSI module in the location case used for the SSI equipment of the new signals (O'Neill 1999, p43). Whichever method is used, the signals obtained are fed into lineside electronic units that contain pre-defined telegrams. Depending on the aspect, the appropriate set of telegrams is then selected by two independent telegram generators and fed through selection logic to a modulator / transmitter connected to the inductive cable loops. These telegrams include specific data about the line ahead, a loop identity code and the distance to the next loop. This provides the train with the data it needs to calculate braking curves whilst avoiding cross talk problems and ensuring detection of a failed loop (Barnard et al. 1992, p2 & 3). The transmitted messages consist of 83 bits, transmitted using Frequency Shift Keying at a carrier of 36 kHz (+/- 400Hz) (Uebel 1996, pC1/13). Telegrams take 280ms to send and are transmitted continuously (Barnard et al. 1992, p3).

Two antennas on the train pick up the transmitted signal, which is used by the Vital On-Board Controller (VOBC) (Uebel 1996, pC1/14-15). In areas with bi-directional running, loops are laid out in both directions and operate at the same frequencies. Each loop is only valid for one direction and the VOBC determines from the sequence of loop numbers which is the correct one for its direction of travel (Barnard et al. 1992, p9). Unlike the GWML implementation, the layout of loops in the Chiltern scheme does not provide any in-built directionality. This must, therefore, be determined by passing at least two loops (Wright et al. 2002, p2).

The VOBC operates on a 2 out of 2 basis. In normal operation, each channel generates a dynamic life signal which is monitored by a supervision board. The train's emergency brake is only released if both processors' life signals are detected. If either processor's internal tests fail, or cross comparison of data shows discrepancies, the life signal is not generated. (Barnard et al. 1992, p7). The VOBC can process up to four targets at a time, selecting the most restrictive and calculating four braking curves for it:

- The first curve is an 'indication' curve, used to determine when target data should be displayed to the driver;
- The second 'warning' curve is used to generate audible alarms to the driver if the train exceeds the maximum safe speed;
- The third 'service brake' curve determines when the service brake should be applied if the train is not brought under control following a warning;
- The fourth 'emergency brake' curve is then used to trigger application of the emergency brake if the service brake fails to operate adequately.

The first three curves end at the signal, but the emergency brake curve ends at the end of the overlap. The emergency brake is also applied automatically if the train passes a signal at danger and can only be released once the train has come to a stand (Uebel 1996, pC1/16).

Fixed train characteristics (such as nominal wheel diameter and braking performance) are compiled into the VOBC software. Other train parameters (such as actual wheel diameter) can be entered during maintenance or by the driver at start up (Barnard et al. 1992, p9). The train then determines its position and speed by use of tacho generators, accelerometers and the antennae (to detect loop transpositions) (Barnard et al. 1992, p2; Uebel 1996, pC1/15).

There is no facility for implementing emergency speed restrictions in the Chiltern ATP system. They are implemented purely on the basis of operating procedures. (Wright et al. 2002, p2).

Four problems encountered during the Chiltern Lines scheme are particularly worthy of note:

- The BR philosophy of prohibiting use of release speed for a train that is actually over a loop caused the drivers problems in aligning the train with CCTV monitors for one person operation where these were located close to a signal at red (the philosophy was later relaxed in order to overcome this problem);
- The class 165 trains on the Chiltern Lines also had a three step brake control which was not reflected within the ATP braking algorithms, forcing drivers to continually move the brake control back and forward between settings to mimic the ATP profile rather than driving normally;
- Whilst problems of insufficient infill are theoretically easy to overcome by extending the existing loops, operating experience has shown that this has not done enough for the actual usage of the route, with resultant performance limitations;
- Chiltern experienced tachometer problems due to the shock and vibration levels caused by the UK infrastructure being well in excess of those seen in Europe, where the same tachometers had previously been used. The tachometer reliability is still an issue and, currently being resolved by detailed analysis of the tacho units during maintenance and change-out where required.

  (Holgate et al. 1993, p11; Wright et al. 2002, p2; Wright 2004, p9).

In contrast to the GWML ATP scheme, Chiltern Railways (the only operating company to use the areas fitted with the Chiltern Lines BR ATP system) have fitted all new rolling stock with the ATP system. They also operate a strict 'no ATP, no go' policy in the areas of line fitted with ATP. The system is, therefore, operating as a true intermittent ATP system (see Appendix A, interview with Tim Brockbank).

C6.2 SWEDISH ATP (JZG 700)

A beacon based intermittent ATP system known as the JZG 700 Automatic Train Control System is in use throughout the Swedish state railway network.

The train is fitted with an antenna, and scans the track continually with a 27MHz signal. When a beacon is passed, it becomes energised by this signal and returns a message at a frequency of 4.5MHz, with a transmission rate of 50kbits/sec. Messages are transmitted in the form of 32 bit telegrams, with at least 8 telegrams being possible from the same transponder, even at speeds of 300km/h. This arrangement means that no local power is required by the transponders (Rose et al.1989, p3).

The beacons are paired, so as to provide high reliability, higher communication capacity and automatic detection of train direction. The first beacon sends a message defining the speed limits and conditions of the track ahead. The second specifies the distance to the next information point and a third transponder can also be used if additional information is required. The beacons can be permanently wired to send the same message, or can be connected to an encoder unit connected to a signal, or other trackside equipment. The received messages are processed by two independent and diverse programmes in a trainborne microcomputer, the result only being accepted if both results match. The microcomputer then uses the train's speed measurement system to calculate a distance to go speed profile, which it continually monitors to ensure that the train speed always remains at a safe level (Rose et al.1989, pp3 & 8).

The driver is given two 3 digit displays:

- A main display of the current speed limit and;
- An auxiliary display of the next target speed.

When the train passes the point where it determines that braking should commence, a short audible warning is given to the driver and the value in the auxiliary display is transferred to the main display, which then flashes. This indicates that the driver has 8 seconds to start braking. Failure to do so causes the audible warning to sound 6 times in 5 seconds, after which the service brake is automatically initiated if it has still not been applied, or has only been applied with insufficient force. If the service brake fails to apply, an emergency brake is then initiated after 2 seconds. If the driver does bring the train speed down to the required speed before intervention, the system continuously monitors the speed to ensure that it remains below the maximum permitted (Rose et al.1989, p8).

When the point to which the target applied is passed, the speed indicator stops flashing (Rose et al.1989, p8).

An immediate emergency brake application is also initiated if a train passes a signal showing a stop aspect (Rose et al.1989, p12).

C6.3 THE 'CROCODILE' PRE-WARNING SYSTEM

When speeds were raised to 200km/h in 1968, SNCF adopted a pre-warning signal (flashing green) to cater for the longer stopping distances. An inductive loop in the track (operating at 6 to 100 kHz) was developed to replace the functionality of the 'crocodile' and provide intermittent transmissions at signal locations. Trainborne systems transfer automatically from 'crocodile' to 'pre-warning' operation on passing an activated pre-warning inductive loop. The pre-warning system then includes all of the information previously transmitted by 'crocodile', together with a speed supervision system (in accordance with a pre-determined speed curve stored in the on-board memory). When a train detects a pre-warning loop, the letter 'B' is illuminated in the driver's cab to advise him/her that he has entered a 160 to 220km/h section of line. When a distant signal is passed displaying a flashing green aspect, the train detects this from the inductive loop and triggers flashing illumination of a 'P' symbol in the cab for 12 seconds, together with the sounding of a bell. The bell continues to sound until an acknowledgement button is pressed by the driver. If the driver fails to keep the train below the pre-determined speed curve (which reduces to 160km/h before passing a second distant signal at yellow), an emergency brake application is initiated and a red display 'URG' is illuminated in the cab (Bailey 1995, pp268-9, 278-9; Guilloux 1991, pp46-7).

C6.4 TRANSMISSION BALISE LOCOMOTIVE 2 (TBL 2)

An advanced version of TBL has also been developed, known as TBL2. This system is beacon based and is broadly similar in functionality to KVB and the Great Western ATP system (Pincock 1998, p242).

TBL2 differs from TBL1 in that it also provides:

- monitoring of braking curves for track occupation and permanent or temporary speed restrictions;
- Supervision of maximum line and train speeds;

  (Bailey 1995, pp270)

Beacon transmissions occur in bursts of 40μs, each of which represents a bit of binary data. If the bit is a '1', the burst occurs at 110kHz. If it is a '0', it occurs at 90kHz. Each byte of data is selected from a predefined list of 11 messages, such that each legitimate byte differs from all other legitimate bytes by at least 4 bits (Pincock 1998, p243). Messages are transmitted continuously with 119 useful bits and redundant coding to ensure correct transmission (Bailey 1995, pp270). There are two varieties of TBL beacon. The first is a 1m long loop of stainless steel bar. The other is a loop of wire that can be up to 100m long (which allows messages to be transmitted over a wider area if required to overcome the latency problems usually associated with an intermittent system). Both types are fitted to the left of the track centre line to allow automatic detection of running direction and are connected to other trackside equipment in order to define the signal to be transmitted (and obtain the required power) (Pincock 1998, p243).

The trains are fitted with two search coils aligned with the edges of beacons that are passed in the correct direction, such that they detect equal and opposite magnetic fields. If a train passes a beacon in the wrong direction, a field is only detected by one coil and consequently ignored by the TBL processor (Pincock 1998, p243)

The trackside TBL2 system is duplicated, as a two out of two system, whilst the trainborne equipment is triplicated, in a two out of three arrangement. The transmitted messages include all data required up to the next beacon, and this data is stored on-board the train. The speed supervision curve is computed on board, accounting for speed restrictions (including the distance to their start location), average gradients for the section ahead, train braking characteristics, train length, line speed and the reported speed from the trains tachometers (Bailey 1995, p271).

C6.5 L 10 000 / EBICAB / CONTROLE DE VITESSE A BALISES (KVB)

L 10 000 and Ebicab are similar and compatible systems in use on BV and NSB since 1978 and 1979 respectively. The Controle de Vitesse a Balises (KVB) system was introduced on SNCF in 1989 as a modified version of the L 10 000 and Ebicab systems, developed in order to allow for quick implementation after a spate of serious accidents on SNCF. As the functionality and physical characteristics of these three systems are very similar, the author will concentrate on a description of KVB, noting variations in the other systems. (Bailey 1995, pp271-2 & 278-81; Guilloux 1991, p49).

The KVB system utilises two types of transmission beacons. The first type transmits fixed data, such as permanent or temporary speed restrictions. The second transmits variable data that is dependent on signal status. In order to achieve the correct status transmission, encoder units are connected in parallel to each signal lamp. These determine the status of each signal aspect and encode this information, which is then relayed to the variable beacons for transmission to passing trains. (Guilloux 1991, p49)

The beacons and train antennas function as both receivers and transmitters, with signals passing in both directions. The two signals operate at different carrier frequencies, in order to avoid interference: track to train at 4.35MHz; train to track at 27.115MHz. The train transmits bursts of signal that last for 17.5μs and are separated by 2.5μs gaps (i.e. a burst frequency of 50kHz). These transmissions do not transmit information, but act as a source of power for the beacon and as a datum to enable the beacon to synchronise its own emissions with the antenna's 50kHz bursts (Pincock 1998, p242).

KVB can transmit eight messages to the train at speeds of up to 300km/h. Four of these must be received identically for the message to be considered as valid. Each message consists of four codes that are each made up of four coding bits and four validation bits. Three of these codes contain data, whilst the fourth is used for synchronisation. The data transmitted to the train confirms the speed limit at the current and next signal, the distance to go between the two signals and the gradient over that distance. In addition to this data, the maximum train speed, train length, maximum deceleration profile on a level gradient and the train type are also stored on the train. The train borne KVB system determines actual speed from the trains tachometers, and performs speed supervision based upon real time calculations (Bailey 1995, pp271-2 & 278-81). This supervision continuously enforces the lowest of the speed limit of the train itself and limits imposed by the track (such as curves, points, work sites, etc.), as well as controlling deceleration on approach to a speed restriction or stop signal (Guilloux 1991, p49).

The SNCF implementation warns the driver if the train over speeds by 5km/h or more, or if the braking distance becomes too short, but leaves service braking within his/her control. The system then ultimately activates the emergency brake if the actual speed exceeds the permitted speed by 10km/h or if the driver does not react to a braking distance warning within 5 seconds. On BV, however, the L10 000 system will activate the service brake itself when an overspeed is detected. (Bailey 1995, pp271-2 & 278-81; Montadert 2000, p950).

On SNCF, the KVB system provides no target speed information to the driver, only stop or proceed indications. The driver is then expected to drive to lineside signal aspects. On NSB and BV, however, the Ebicab / L10 000 systems are considered to be fail safe, and more elaborate speed information is provided in the driving cab. The driver is then expected to use lineside signals as a fall back should the ATP fail. (Bailey 1995, pp271-2 & 278-81). In addition to the normal warning indications, the cab display panel can also provide information obtained from received data for the purposes of driver assistance. (Guilloux 1991, p49).

In accordance with the different applications, SNCF operating rules impose a maximum speed of 160km/h on train movements if the on-board KVB system fails, whilst trains are restricted to 80km/h on BV in the same circumstances (Bailey 1995, pp271-2 & 278-81).

During the first 10 years of operation on SNCF, the number of overspeeding trains and near missed were found to decrease, with equivalent network constituency, traffic levels, numbers of rolling stock and drivers. However, the intermittent nature of the KVB transmissions was also found to affect suburban line capacity. Perhaps more disturbingly, the system was also found to have the unintentional effect of changing driver behaviour: audits showed that some drivers began to use the KVB display to guide their driving, focusing on following the 'safe approach speed' rather than targeting a stop at a red signal (Montadert 2000, pp952-3).

C6.6 ZUB 100

The ZUB 100 system uses beacons for intermittent transmission or cable loops for continuous transmission. Beacons, located by lineside signals, transmit a combination of fixed route data (stored in a chip within the beacon) and signal aspect information (obtained from the lamp circuit or via a combination of contacts). The beacons are powered by induction at 100kHz from passing trains and transmit their signal in both a normal and inverted form via independent channels at 50kHz and 850kHz. The transmission is only considered to be valid if two consecutively received telegrams have the same content (Geduhn 1995, p48).

In-fill beacons or loops can also be installed on the approach to a signal for anticipation of signal aspect changes (Geduhn 1995, pp48 & 50).

Two identical trainborne computers calculate a brake monitoring curve based on train data and the received route data. The target speed is then taken from this and indicated to the driver as a red pointer on his/her speedometer. The permitted speed at the next signal is displayed by a 5 digit display and the clear track distance to the next signal with a stop aspect is also shown as an illuminated bar alongside the speedometer. This information is updated quasi-continuously, based on the last received transmission and adjusted to account for subsequent movement as detected by the train's odometry system (Geduhn 1995, p49).

If the train's current speed reaches the maximum permissible speed, an alarm tone is sounded in the driver's cab. If the current speed then exceeds the maximum permissible speed, an automatic service brake application is initiated (subsequently followed by an emergency brake application, if required) (Geduhn 1995, p49).

The system allows a signal displaying a stop aspect to be approached at a release speed. If the train passes the signal whilst being supervised against the release speed, an automatic emergency brake application occurs and is enforced until the operation of a release button after the train comes to a stand (Geduhn 1995, p49).

Information can also be transmitted from the train to the trackside by reversing the transmission direction of the train's coupling coil and installing a receiver antenna in the track (Geduhn 1995, p48).

In its continuous form, ZUB 100 uses LZB trackside equipment (Geduhn 1995, p48). This configuration will not, therefore, be discussed further here (see section C7.4).

C6.7 ETCS LEVEL 1

The specifications for the European Train Control System (ETCS) has been developed by UNISIG, an umbrella group for all of the major European signalling suppliers. In order to provide for the needs of different railways (both high speed and other types of main line), the system has been specified in three levels of implementation, known as levels 1, 2 and 3 (Pore 1996, p51). The specifications are intended to enable interoperability, both between the products of different suppliers and between different train service providers. The Level 1 system (which offers intermittent ATP) will be discussed in this section, whilst the level 2 and 3 systems (which offer continuous ATP) will be discussed in sections C7.10 and C7.11 respectively.

The ETCS level 1 system is currently being tested across Europe, including at Old Dalby in the UK. However, it is not likely that it will be ready for deployment in the UK (beyond limited trials) before 2003 (Muttram 2001, p6).

The level 1 system is designed as an overlay on a conventional signalling system, with conventional train detection and train separation functions being performed by the trackside equipment of the underlying signalling system (track circuits, interlocking, etc.) (Thomas et al 1992, p28). Level 1 operates by use of intermittent 'spot' transmissions from Eurobalises (beacons that transmit information in accordance with a standardised European specification, each with a transmission range covering approximately 1m of the track). These balises are powered by electromagnetic induction from the antenna of a passing train. The receipt of a telepowering signal (which is in the range of 27.095MHz to 27.115MHz) causes a balise to send a reply telegram at a carrier frequency of 4.24MHz, modulated by Frequency Shift Keying at 282.2kHz. By this means a telegram of either 1023 or 341 bits can be transmitted (including 830 or 210 usable bits) at a data transmission rate of 564.48Kbits/s. This rate enables the balise transmission system to operate at speeds of up to 500km/h. If one balise is not sufficient to send the required message, balises can be grouped so that their combined telegrams form the complete message (Lundberg 2002, pp1,3-5; Reddy 2001; UNISIG 1999, chapter 2 p6).

Two types of balises can be used in Level 1:

- Fixed balises - designed to transmit a default telegram when a telepowering signal is received. This telegram is stored internally in Non-Volatile memory
- Switchable balises - designed to transmit telegrams that are supplied to them by Lineside Encoder Units (LEUs), so that the message can be changed in accordance with the state of the lineside signalling

  (Reddy 2001).

As a minimum, switchable balises are located at every stop signal in a fitted area. Additional switchable balises can also be used to provide infill information on the approach to the signal, where required. This is typically done to reduce the delays that can be caused by the ATP system enforcing a braking curve even though the signal being approached has actually cleared to a less restrictive aspect. The messages from switchable balises include details of movement authorities and track description data for the route to be taken (UNISIG 1999, p8).

Fixed balises can be used to provide information such as a location reference, details of permanent speed restrictions and track geometry that will not change with route conditions. In order to ensure safety, the location accuracy is within 1m for each balise passed (Lundberg 2002, p3).

The ETCS system requirements specification also allows for the use of track loops (known as Euroloops) and radio infill to provide more continuous information updates on the approach to a signal. With semi-continuous infill provided, it would theoretically be possible to remove the lineside signals and operate a level 1 system on the basis of in-cab signalling (UNISIG1999, chapter 2 pp7 & 17). However, it is not currently planned to utilise these forms of infill transmission for UK applications.

LEUs are required to interpret the state of the underlying signalling system and to determine the messages that should be transmitted to trains by switchable balises. As the underlying signalling is not of the same type (or even operating to the same principles) for all of the railways that will implement level 1 systems, the way in which this is done is railway specific. In the UK, current sensing devices known as TCS LITs will be connected to signal heads in order to determine the aspect that is being displayed. The LEUs will then utilise this information to select the appropriate messages for transmission to trains.

The trainborne ETCS system receives absolute location information from balise messages. It then uses a trainborne odometry system to determine the trains speed and location beyond the last received balise location (Woodland[1] 2000, p5; ERTMS Users Group 2001, p8). Based on this information, together with the content of received balise messages and the train characteristics data that is stored on-board, the trainborne system provides continuous speed supervision and protects against overrun of the movement authority within a level 1 fitted area (UNISIG 1999, p17). In an area of track not fitted with trackside ETCS equipment, the trainborne equipment still provides supervision, but only against a ceiling speed for the train type (Woodland[1] 2000, p6).

If the movement authority is exceeded, the trainborne system initiates an emergency brake application. To recover from this, the train driver must acknowledge the intervention and the train must be brought to rest (UNISIG 1999, pp32-3).

The trainborne system also provides in-cab information to the driver. However, the intermittent nature of the track to train communications makes it likely that there will be occasions when the information provided to the driver by the Level 1 in-cab display does not agree with the information that can be obtained from the lineside signalling. The potential for discrepancies between lineside and in- cab information will be minimised where the location of the intermittent transmissions is optimised for the expected train performance and scheduling. It has, therefore, been noted that a level 1 type system is not ideally suited to mixed traffic lines where the performance differs between trains (Holtzer 1999, p588).

In order to mitigate this problem, it is the intended UK operating practice that the driver should use the most restrictive of the information sources when there is a discrepancy. Where the level 1 system is the most restrictive source, this practice will be enforced by the systems supervision (which may prove frustrating for drivers and will have an impact upon performance). However, where the underlying signalling is the most restrictive, reliance will be placed upon the drivers observation and application of the correct practice. In order to reduce the likelihood of confusion, the in-cab display will indicate speed and distance data to represent the envelope that the level 1 system is supervising against, but the current movement authority will not be displayed. The level 1 system will also provide a warning when the supervised speed is exceeded by a defined margin. The driver should then reduce speed. If, however, he/she fails to do so sufficiently to prevent the train from exceeding the supervised speed by a second (larger) margin, an alarm indication and service brake application will be initiated. This application will only be cancelled when the speed falls to (or below) the supervised value and the driver acknowledges the intervention. Whilst this arrangement should not present difficulties in most situations, there is the potential for the driver to be misled by the in-cab display (particularly when the ETCS in-cab information fits the driver's expectation pattern – such as when a signal has been replaced to danger in front of an approaching train) (Bott 2000, pp15 & 17).

As an overlay system, Level 1 will, unfortunately, reduce the available capacity of any railway to which it is fitted. Whilst this capacity reduction can be minimised by appropriate use of in-fill transmissions, it will still make an appreciable difference to any railway attempting to achieve high capacity throughputs and rapid recovery from disruption. The overall effect of immediate implementation of a level 1 system onto the UK Mainline Railway was modeled by the ERTMS Programme Team and found to be a capacity reduction of:

- 15% of network capacity without infill
- 11% with a single infill balise per signal

(Waboso 2002, p47)

C7 CONTINUOUS ATP SYSTEMS

In section 4.3.2, Continuous ATP was defined as the second sub-category of Comprehensive ATP. It is therefore also 'a system automatically enforcing compliance with **all** speed restrictions **and** movement authorities (for all vehicles) within a given area'. However, continuous ATP offers a more advanced solution compared to intermittent systems, with the trainborne system able to update the supervision conditions at much more frequent intervals along the railway. This makes continuous ATP systems much less restrictive to the railways capacity than an equivalent intermittent system.

C7.1 AUTOMATISCHE BEINVLOEDING (ATB)

ATB, in use on NS since 1966, relies on one-way (track to train) transmission via coded track circuits. These transmit one of five speed codes as amplitude modulation on a 75Hz carrier signal (for 40, 60, 80, 130 and 140 km/h). In practice, the five codes are not sufficient to describe accurately the speed limit for all sections of track. Where the actual speed limit is not available as a code, the transmitted code and subsequent in-cab indication are selected to be the nearest speed code above the actual speed limit. The in-cab indication is therefore seen as a guide only, with lineside signs and signals being the authoritative source of driving information.

Whenever the speed information changes, an audible warning (gong) is given in the cab. The trainborne system then supervises the train's speed against that given by the code. If the train is detected as overspeeding without a brake application in progress, another warning (this time a bell) is sounded. If the driver does not then activate the brakes within a few seconds the emergency brake is automatically applied (and remains applied until the train comes to a stand).

There is no code for stop, the meaning of no code present being 40km/h. This means that the system is only able to provide any protection at speeds higher than 40km/h, responsibility for safety below that speed remaining with the driver. The sounding of a buzzer every 20 seconds whilst the train is in receipt of a 40km/h code and the brakes are not being applied assists him in this. The buzzer acts as a vigilance reminder and can be cancelled either by operation of an acknowledgement button or by application of the brakes. One advantage of this approach is that it provides graceful degradation, allowing continued train movements at 40km/h on a line of sight basis in the event of signalling system failure (Bailey 1995, pp273-4 & 280-1).

The use of target speed code transmission through track circuits to determine supervision limits makes the ATB system best suited to railway operation involving trains that all have similar performance profiles. Where this approach is used with train profiles that differ significantly the infrastructure enforces a standard speed profile on all trains, reducing system utilisation (Fenner 2002, p32).

C7.2 TRANSMISSION VOIE MACHINE 300 (TVM 300)

With the introduction of TGV trains on SNCF, running at speeds of 220km/h, it was considered that lineside signals were better avoided. TVM 300 was therefore developed to provide in-cab signalling that would support the flexibility and adaptability of driver control whilst providing a speed supervision system (Guilloux 1991, p47). It was introduced in 1981, based on coded track circuits, modulated at low frequencies between 0.88 and 17.52 Hz (Bailey 1995, p273).

Under this system the line is divided into 2100m blocks (on flat sections), within which coded frequency messages consisting of up to 18 data items are transmitted. The messages relate to the status of the next block and the location of any points that need to be passed at lower speeds, with status being obtained from the interlocking. The end of each block section is identified by a trackside marker board (Guilloux 1991, p47; Pincock 1998, p244).

The transmitted speed information and subsequent speed supervision are stepped (with only specific values being valid) and include use of an overlap track. A representation of the braking curve that results from this is shown in Figure 7-1.

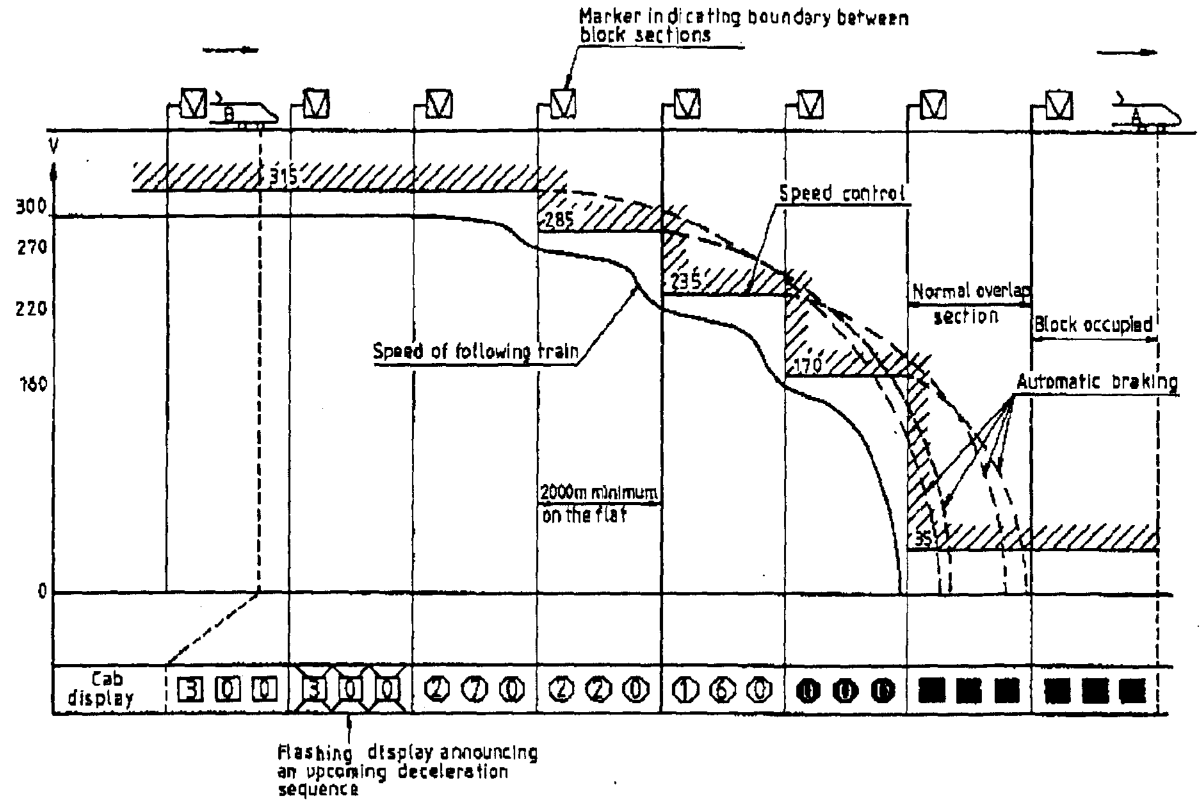


Figure C 7-1: TVM 300 Braking Curve Arrangement (Guilloux 1991, p52)

The electronic systems on the train use the transmitted information to display authorised and target speeds to the driver (interpreting the speed codes by means of look up tables for the train type) and to supervise the trains' actual speed against these speeds. The emergency brake is automatically applied if the train exceeds the speed allowed by more than 5 to 15km/h (depending on the speed) or overruns a 'do not proceed' marker (Bailey 1995, p273 & 278-9; Guilloux 1991, p47; Moens et al. 2003, p8).

There are three areas to the driver's display panel, each of which can show all of the valid speed codes. The first area, 'Vitesse Limite (VL)', shows the maximum line/train speed limit. The second, 'vitesse d'Annonce (A)', shows the target speed at the end of the section and the third, 'vitesse d'execution (E)', shows any speed limit due to restrictions. In addition to these codes, a display of three red squares indicates an immediate stop, either due to an emergency instruction from the control centre or because the train has overrun a section with a zero speed code. This indication is also given if the continuous signal from the rails is lost, cannot be understood or is of the wrong frequency, if the driver selects reverse or if the driver selects the self-test function whilst TVM is active (Pincock pp245-6).

Where it is not installed on the track (and hence there is no continuous signal in the running rails) the trainborne TVM system must be 'disarmed' to avoid default to immediate stop. This is achieved in two ways. The driver can operate a switch in the driving cab, or it can occur automatically on receipt of a special message from a beacon located at the end of the TVM zone. Arming on entry to a TVM zone is also achieved in these two ways (Pincock 1998, p248).

TVM 300 can be used at speeds of up to 300km/h (Bailey 1995, p273 & 278-9; Guilloux 1991, p48).

The TVM300 system offers a high level of safety, but suffer from the need to invest heavily in infrastructure based equipment. It also requires high levels of maintenance activity to operate reliably and safely (Schmid 1999, p3).

As with ATB, supervision based on target speed codes transmitted through track circuits makes the TVM 300 system best suited to railway operation involving trains that all have similar performance profiles, where the codes used can be optimised to that profile (Fenner 2002, p32).

C7.3 TRANSMISSION VOIE MACHINE 430 (TVM 430)

TVM430 was introduced in 1993 and is used on SNCF and Eurotunnel. This system is compatible with TVM300 and operates with the same basic functionality and physical characteristics. The main differences are that:

- The TVM430 coded track circuit signals are capable of transmitting messages with 21 useful bits and 6 coding bits;
- The trainborne system is based on a duplicated microprocessor;
- The available codes are used to transmit an adjusted set of speed graduations that include the length of the current track section, permissible speeds in the current and next track section, mean gradients and operational instructions;
- The trainborne system decodes the received messages several times per second and uses the data that they contain to calculate supervision curves;
- A flashing indication is given in the driver's cab to provide advanced warning that the next section will enforce a more restrictive speed;
- The revised speed graduations and advanced warning of speed reductions mean that TVM 430 can operate with 1500m blocks on (flat sections). For the lower speeds of the Channel Tunnel this is reduced further to 500m;
- The train's actual speed is supervised against the calculated curve and the emergency brake is applied if an overspeed is detected (i.e. whereas TVM 300 operates purely on the basis of a series of target speeds, TVM 430 also utilises the distance to go to a target in order to provide a more refined supervision).

  (Bailey 1995, pp273; Bin 1996, p114; Guilloux 1991, p48; Pincock 1998, p244).

A representation of the braking curve that results from this is shown in Figure 7-2.

TVM 430 can be used at speeds of up to 350km/h (Bailey 1995, pp 278-9). Train speed and distance travelled are measured in triplicate, each measurement being based on two different data sources and independently corrected to allow for slip/slide. The ATP system then uses all three measurements in its calculation of stopping sequence. By this approach, speed accuracy to +-2% is achieved (Systra 1999,pp14 & 20)

Whilst the first TVM 430 installations utilised a cab signalling/ATP controller that obtained information concerning the state of the railway from a separate interlocking, the Mediterranean High Speed Line and Channel Tunnel Rail Link (CTRL) installations have utilised a combined interlocking and cab-signalling control system. In France, this is known as SEI (Systeme d'Enclenchment Integre), whilst in the UK it is referred to as ITCS (Interlocking Train Control System). These systems concentrate track circuit transmitters and receivers at the central interlocking, which effectively limits the area of control to 15km of line per system (the maximum distance over which the track circuits can be fed being 7.5km) (Moens et al. 2003, p3).

In its original form for use by SNCF, TVM 430 only supported operation of high-speed passenger traffic. For the CTRL implementation in the UK, modifications have been made (through addition of extra track codes and trainborne look up tables) in order to support mixed traffic operation. However, it has not been possible to give a maximum speed display in the cab corresponding to limits of freight traffic, only a flashing display of three green squares. The maximum speed is then left to the freight operator's train and route knowledge. The fixed train class look up tables used by TVM when determining braking curves also limit the system to supervising against a standard set of tran characteristics, with no means of accounting for variations in performance with different loading, etc. The system is, therefore, still not ideally suited to freight operation (Moens et al. 2003, pp1, 8).

TVM430 allows for a data acquisition and display time of 2.5 seconds, coupled with a driver reaction time of 5 seconds (Systra 1991, p12).

As with TVM300, the TVM430 system suffers from the need to invest heavily in infrastructure based equipment and requires high levels of maintenance activity to operate reliably and safely (Schmid 1999, p3). Whilst more optimal control of train speed can be achieved during braking to a target speed code than under TVM430, the fixed track circuits transmitting the codes still mean that supervision can only be truly optimised for one train performance profile. TVM430 is, therefore, still best suited to operation of services with similar train performance characteristics.

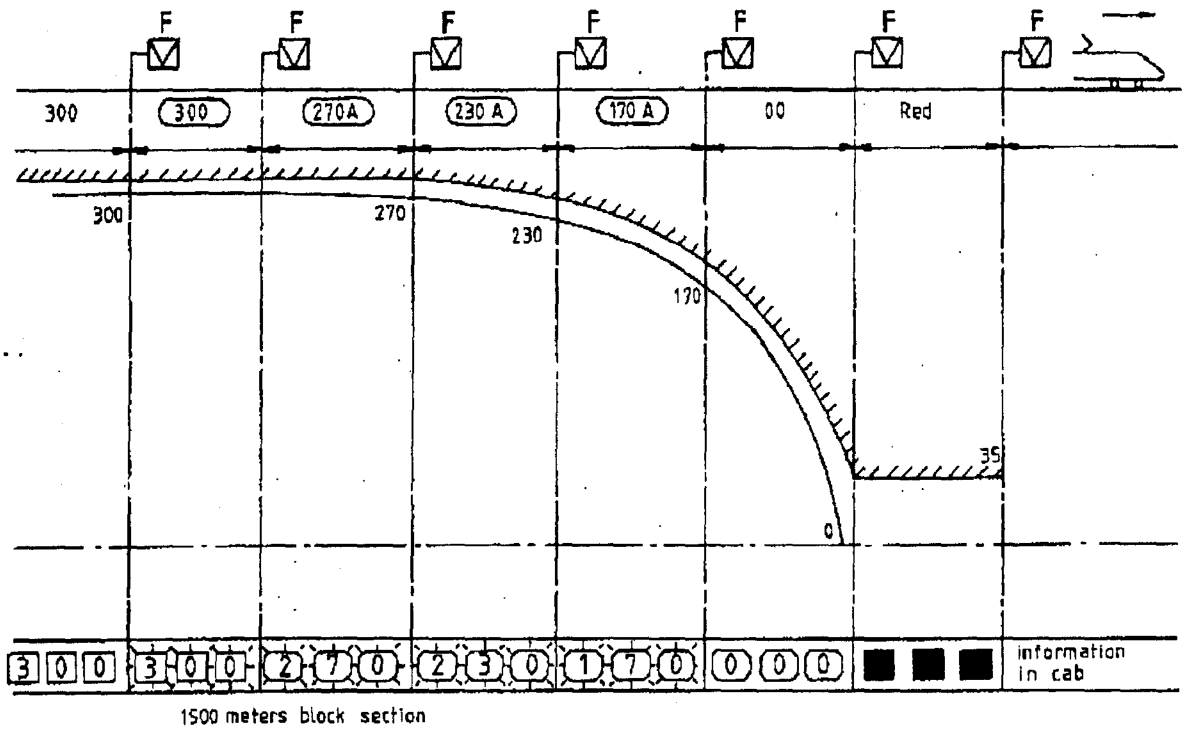


Figure C 7-2: TVM 430 Braking Curve Arrangement (Guilloux 1991, p53)

C7.4 LINIENZUGBEEINFLUSSUNG (LZB)

LZB is now in use on ÖBB, DB AG and Renfe. It was first introduced in 1965 based on a hardware solution and then upgraded in 1972 to a computer-based system (Bailey 1995, pp276-7; Uebel 1996, pC1/8). LZB utilises full duplex transmission through continuous loops laid in the track. Each conductor can be up to 12.7km in length (with transpositions every 100m to aid in determining train position) and is connected to a centralised line centre (operating in a 2 out of 3 computer arrangement). Each line centre monitors up to 16 loops and is also connected to all interlockings in the area, as well as to adjacent line centres. The line centres are also provided with operator terminals for accessing information and applying controls (such as speed restrictions) (Bailey 1995, pp272; Uebel 1996,ppC1/8-10).

On entry to an LZB section, each train transmits its characteristics (such as braking performance, length and category), based on data input on-board. It then continuously repeats transmission of its location, actual speed, braking characteristics and internal operational data. This train to track communication is achieved by 41 bit telegrams (28 of which are useful) transmitted by means of 200Hz modulation of a 56kHz carrier (Bailey 1995, p272 & 276-7). This transmission occurs at a rate of 600 bits per second. (Uebel 1996,pC1/11).

All fixed data (speed restrictions and geography) is stored in the LZB control centre. The interlockings forward signal aspects, point settings, level crossing barrier positions and emergency stop commands to the LZB control centre, which returns level crossing approach messages, signal setting commands, operating state and fault data to the interlockings. The control centre then uses the data available to it to determine targets (such as distance to stopping point). These can be greater than would be permitted by the lineside signalling alone, since the control centre is aware of several block sections ahead. This data, together with other operating information (including speed and distance targets for the driver) is then transmitted to each train at least once every second (Bailey 1995, p272; Uebel 1996, ppC1/8 & 11). The train characteristics set is flexible enough to support mixed traffic operation on an LZB fitted line.

Track to train communication is achieved by 83 bit telegrams (67 of which are useful) transmitted by means of 400Hz modulation of a 36kHz carrier (Bailey 1995, p272 & 276-7). The transmission occurs at a rate of 1200 bits per second (Uebel 1996,pC1/11).

The trainborne computer (operating in a 2 out of 3 arrangement) utilises the received messages and tacho intputs to calculate the maximum permissible speed, control indications to the driver and monitor train speed. In order to achieve this reliably, the trainborne system receivers and tacho-generators are duplicated for availability.

The driver is given indications of actual speed and target speed (in the form of yellow and red needles respectively on an analogue speedometer). The distance to go before reaching the point at which the target speed applies is also shown alongside this. If the permissible speed is exceeded by more than a permissible margin an emergency brake application is automatically initiated (Uebel 1996,ppC1/9 & 28). Absence of telegrams also causes automatic application of the trains brakes (Bailey 1995, p272 & 276-7).

The LZB system is applied on lines with speeds of up to 300 km/h. 'The ATP function is complemented by a number of other functions that share the same data and processing facilities, such as basic automatic driving and in-cab signalling' (Thomas 2001, p112). Where lines are used only by fitted trains, lineside signals are not required. They are, however, retained for use by unfitted trains (albeit in reduced numbers on new lines) (Bailey 1995, p272 & 276-7).

LZB is also considered to be a typical example of a technology that offers a high level of safety, but suffers from the need to invest heavily in infrastructure based equipment and requires a high level of maintenance activity (particularly related to track loops) to operate reliably and safely (Schmid 1999, p3).

C7.5 FS2000 ATP SYSTEM

The FS2000 ATP system is currently in operation on LUL's Central Line and on the Singapore Mass Rapid Transit system. It is a fixed block, track circuit based, system utilising FS2500 jointless track circuits (Jeffrey 1999, p1; Taskin et al. 1995, pA3/15).

The trackside system utilises 14 ATP codes to provide 7 vital ATP speed bands. These are applied to track sections so as to maintain a safe separation between trains and other obstructions, and to enforce speed restrictions. They are transmitted to the train through the coded jointless track circuits, which operate at 8 different carrier frequencies (such that adjacent track sections always operate at a different frequency to avoid problems of read through) (Jeffrey 1999, p2). The track circuit carrier frequencies used are between 4080Hz and 6000Hz. The ATP codes then use modulation at frequencies of 28Hz to 80 Hz (see Appendix A, interview with Steve Rodgers).

For the track circuit to register the absence of any train, it must receive an FSK signal that contains both the correct carrier frequency and a valid vital ATP code modulation frequency. This arrangement provides a high level of immunity to traction interference (Jeffrey 1999, p3).

The ATP code is also detected by the train and is used to determine the permissible speed in both the current block and the block ahead. The train's ATP system uses this information to supervise against the permissible speed for the current block. If an overspeed is detected, the emergency brake is automatically applied and maintained until the speed returns below the permissible speed. The train is therefore only brought to rest following an intervention if the permissible speed was 0mph. In addition to its speed supervision function, the ATP also provides the train driver / ATO system with a target speed for the next block (displayed within the cab, integrated into the speedometer unit). This allows the driver / ATO system to control the train by use of the normal service brakes, such that it enters the next block below the maximum safe speed and avoids an ATP emergency brake intervention. An audible warning sounds in the driver's cab when the target speed changes (with different sounds for upwards and downward changes), but there is no overspeed warning prior to ATP intervention (see Appendix A, interview with Steve Rodgers).

The current-block speed element of the ATP code is therefore used for the ATP safety functions to maintain a safe separation between trains, whilst the next block element is not used as a primary safety function, but for operational purposes (Jeffrey 1999, p3).

Where the 14 codes available do not provide sufficient combinations of ATP speed bands and target speeds, spot loops are used to transmit an additional 5 frequencies from 6240Hz to 6840Hz. Detection of one of these frequencies causes the trainborne system to apply a different interpretation to the track circuit based signal that it is receiving. As a safety principle, the spot loop frequencies only change the target speed interpretation for the next block, not the most restrictive speed information for the current block. Whilst the track circuit code is received continuously, the spot loop codes are intermittent. The trainborne system therefore retains the last received spot loop code in its memory and applies it to the interpretation of speed codes until such time as another spot loop is passed (Jeffrey 1999, p3-p4; Appendix A, interview with Steve Rodgers).

The train is fitted with ATP antennas on the leading bogie, ahead of the first wheel set. These detect the transmitted frequency codes. Tacho-generators mounted on independent wheel sets of the leading car then provide speed measurement (Jeffrey 1999, p4).

The ATP controller contains a safety unit, designed to SIL4 and a non-vital unit that implements the same functionality by diverse means (Jeffrey 1999, p4). Both monitor the track code and the train's speed and either can apply the emergency brake. Inside the ATP controller, the functionality of each unit is independent. However, outside of the ATP Controller box itself they share components such as the tacho generators and antennas (see Appendix A, interview with Steve Rodgers).

C7.6 SYSTEME D'AIDE A LA CONDUITE L'EXPLOITATION ET LA MAINTENANCE (SACEM)

SACEM is the name applied to a generic ATP system that has been implemented in different forms around the world. Within SACEM, safety-decisions are made on-board individual trains on the basis of information transmitted to the train by trackside equipment, or stored in trainborne memory (Guieu et al. 1995, pp15-17). The common features of all SACEM implementations include:

- The division of the line into sectors, each managed by a SACEM sector computer (typically serving 1 to 3 station areas). The main function of the sector computer is to acquire interlocking data (track occupancy, point positions, etc.) and to transmit this, along with fixed infrastructure information (gradient profile, speed restrictions, etc.), to trains in the sector;
- The division of sectors into sections for transmission of data;
- Indiscriminate transmission of data concerning each transmission section to all trains within that section;
- Detection of transmitted data by trainborne pick up coils;
- Selection of relevant data by each train;
- Determination by each train of its location, based on: detection of beacons in the track; use of a movement detector coupled to a train axle to measure displacement; use of a tachometer coupled to a different train axle to detect a locked movement detector axle and received data;
- Calculation by the trainborne computer of supervision curves that the driver must respect so as not to exceed a ceiling speed, speed restriction or safe movement limit. These calculations are based on the received data, current train speed & location, trainborne device status (including which cab is in use) and additional parameters of the train that are programmed into the trainborne computer (maximum speed, braking rate, weight, length, etc.);
- Control of an in-cab signalling display by the trainborne computer, based on the received data and trainborne calculations. This display acts as a substitute for lineside signalling;

- Provision of speed supervision based on the calculated supervision curves. If an overspeed is detected the emergency brake is automatically applied (by a safety critical output to the emergency brake system) and maintained until the train comes to a stand.

  (Galivel et al. 1991, pp 71, 77, 81-2; Guieu et al. 1995, pp15-17; Wei 1999, p383-5; Appendix A interview with Gilles Poitrasson-Riviere).

Published literature concerning the SACEM system notes that the features outlined above allow trackside processing functions to be performed completely independently of the rolling stock characteristics. This means that mixed traffic types can be handled easily and that the trackside equipment requires no modification if changes are made to the trains operating over it (Galivel et al. 1991, p76; Guieu et al. 1995, p16). However, whilst this is true for most data, there are some exceptions due to the limited processing capacity of the trainborne computer. In order to avoid system delays, some pre-processing of energy calculations is carried out trackside, with the results for each train type using the line then being included in transmissions to all trains. This means that changes in train performance (or the introduction of new train types) would require some limited trackside data changes (see Appendix A, interview with Gilles Poitrasson-Riviere).

The transmission of all required trackside data to all trains does mean that a SACEM fitted train can operate on any fitted section of track without needing to have track characteristics downloaded in advance. Similarly, following modifications to track layouts only the trackside equipment needs to be modified, not the trainborne equipment (Guieu et al. 1995, p16).

The generic SACEM design can be easily adapted to different interlocking and transmission technologies (or to changes in trainborne or sector computers) and supports ATO, when required (Galivel et al. 1991, p84). It also has inherent fall back modes, minimising the disturbance caused to traffic by failure of the SACEM system (Galivel et al. 1991, p83-4).

The adaptability of the generic system can best be seen by considering a few of the diverse implementations that are in use around the world. For the sake of clarity, the implementation used for the Paris RER Line A will be considered first. The main differences for the implementations in Hong Kong and Singapore will then also be discussed.

C7.6.1. SACEM ON PARIS RER LINE A:

On the RER Line A in Paris, SACEM operates in conjunction with lineside signals. Each SACEM sector is divided into sections, in each of which the messages to be transmitted to trains are simultaneously injected into corresponding track circuits. The optimal number and position of these track circuits is determined by the operational performance required, the train performance, cost and response time of the driver. In practice this results in 4 or 5 blocks in a 225m platform area (too few sections limits the achievable headway whilst too many sections would cause frequent changes to the cab signalling on approach to a station and thus prevent the driver from paying adequate attention to what is happening outside of the cab) (Galivel et al. 1991, pp 71 & 77).

Trains leave the depot in 'CML' mode (usually refered to as 'standby' mode as an English translation). Transfer to full supervision then occurs on entry to a SACEM equipped area, without the need for driver intervention. When leaving a SACEM equipped area, the in-cab display provides an indication consistent with the boundary signal until the train is several metres before the signal. The in-cab display is then extinguished and speed supervision ends (Galivel et al. 1991, pp71 & 74; Appendix A interview with Gilles Poitrasson-Riviere).

When a train enters a section with its continuous speed monitoring running and cab signalling switched on, the signal at the end of the section is cancelled (the signal aspects are extinguished and a Saint Andrew's Cross of white aspects on a black background is illuminated near to the signal). This cancellation is revoked when the train passes the signal. The operation of this facility is arranged so that the driver never sees a signal extinguished just in front of him/her. (Galivel et al. 1991, pp71 & 76).

If a SACEM sector failure is detected, the signals within that area all become illuminated again. The signals on entry to the area automatically become SACEM border signals. Trains entering the area then automatically go through the process of leaving a SACEM area as they enter the failed sector, and of entering a SACEM area when they leave the failed sector (Galivel et al. 1991, p74).

There are three types of transmission used by the SACEM system, all of which are designed to be superimposed on the lineside signalling:

- The continuous and indiscriminate transmission of data concerning each SACEM sector (including in this implementation the absolute position of all stop signals, gradient profiles, maximum speed limits, location of stations, correct side door enable information, signal aspects, track circuit states and point positions). These are sent in the form of repetitive messages every 0.3 seconds in the 20kHz to 80kHz range (with 5kHz gaps between frequencies used);
- Semi-continuous train to track transmission to cancel lineside signal indications and send maintenance messages. This transmission works in the last 80m of each track circuit (again within the rails) at a frequency of 140.8kHz;
- Intermittent transmission from beacons to trains sending three types of message. The first is initialisation, by two 3m long beacons that provide an absolute position and enable the train to calibrate its tachometer system by measuring the distance between the two beacons. These beacons are powered by trackside equipment. The second type is relocation, achieved by 2.4m long beacons at up to 500m separation. These are powered by the passing trains. The third type are exit beacons, which are also fed by the passing trains.

  (Galivel et al. 1991, p80-1)

None of the transmission equipment used is fail safe. However, the beacons are designed to guarantee no cross talk and the transmitted messages include a fail safe code to ensure secure transmission. The trainborne computer itself operates using a single hardware and single software channel, with safety being provided by repeated checks (Galivel et al. 1991, pp 80 & 82; Appendix A interview with Gilles Poitrasson-Riviere).

The SACEM cab signalling unit is located above the speedometer and consists of:

- A three digit display that indicates speeds in steps of 5km/h (the left digit being extinguished for speeds below 100km/h);
- A frame around the three digit display that can show a yellow aspect;
- A second frame around the yellow aspect frame that can display a green aspect;
- A yellow indicator;
- A red indicator (capable of displaying a steady or flashing aspect. The flashing aspect shows when the override button is operated when the train is stopped and the red indicator is showing);
- A triangular indicator with an exclamation mark that displays a red aspect in the case of an overspeed.

  (Galivel et al. 1991, p71).

The cab display shows either the current maximum speed (within the green frame indication) or a target speed (within the yellow frame indication). The yellow and red indicators are illuminated to reflect an approach to a stop indication, only one being shown at a time. Restrictive indication changes (other than a speed change in the green frame) are accompanied by an approximately 0.75 second audible signal. A further audible warning is given to the driver when the train exceeds the indicated maximum speed. A tolerance margin of about 10% is then allowed between the target speed and an automatic brake intervention. The overspeed indicator is illuminated after an emergency brake application has been triggered and is only cleared once the train has stopped and an overrun push button has been activated (Galivel et al. 1991, pp71, 74, 82 & 85).

C7.6.2. SACEM ON HONG KONG MTRC:

On the Hong Kong MTRC, the SACEM implementation includes a number of modifications:

- A loop based transmission system has been used for the continuous transmission to/from trains, in order to overcome EMC difficulties with transmission through the rails;
- Emergency stop areas have been defined where trains have to receive permanent authorisation to run (in support of Passenger Emergency Plunger systems);
- No unbraked axles are available, so a specific slip/slide algorithm has been added to the trainborne functions;
- The system has been designed to include ATO;
- In accordance with the intended use of ATO as the primary driving mode, there are no automatic signals (although controlled signals are retained in junction areas);
- The implementation includes facilities to detect track circuit and transmission system failures and automatically adapt the system configuration in order to allow continued operation (albeit at a reduced capacity).

  (Guieu et al. 1995, p18; Appendix A interview with Gilles Poitrasson-Riviere)

Track circuits are still used to detect track occupancy, which is translated into stopping point states that can not be passed by a train when they are restrictive (Guieu et al. 1995, p15-16).

C7.6.3. SACEM ON SINGAPORE MRT:

On the Singapore MRT, safe train separation is achieved through moving block operation (Wei 1999, p383).

The trackside ATP system is based on computers located at main stations in a 2 out of 3 architecture. They utilise block occupancy information from the interlocking (as a fall back) and position reports from trains (as a primary source) to estimate maximum train head and minimum train tail positions. All relevant data about train positions and track status are then transmitted to all trains running within the area (Wei 1999, p383).

Continuous bi-directional track to train transmission in this implementation is achieved via a waveguide. This is used for vital data that allows the train to identify stopping points and speed restrictions as well as non-vital data such as positive train identification and departure / arrival times. The waveguide operates on a frequency of 2.4 GHz using direct sequence spread spectrum transmissions. It is a 0.2cm thick hollow rectangular extrusion installed along the side of the track, of approximately 5 cm high and 10 cm wide. Narrow slots are stamped into its upper surface at intervals of 6.1cm, in order to produce a uniform radiated field at the frequency used. (Wei 1999, p384-5).

In order to determine its position, each train counts the waveguide transmission slots in addition to the usual trainborne odometry systems and beacon updates. The combination of these methods allows each train to send a periodic position report to the trackside system that is accurate to within 4cm (Wei 1999,p383-5).

Intermittent beacons (based on Eurobalise technology) are also used to:

- Initialise the trainborne system at the stabling tracks;
- Initialise the trainborne system on entry to a particular ATC sector;
- Provide precise stopping markers in stations.

  (Wei 1999,p385)

Like the Hong Kong MTRC, the Singapore MRT control system includes ATO. The reliance placed upon this primary driving mode has, however, been taken further such that there are no lineside signals (whether automatic or controlled) provided in conjunction with the system (see Appendix A interview with Gilles Poitrasson-Riviere).

C7.7 SELTRACK

The SELTRACK ATP system was first introduced in 1971 in Europe, and is also in use in North America. It controls trains at speeds up to 200km/h by use of inductive loops with transpositions at regular intervals, as already discussed for the Docklands Light Railway implementation in section 4.2.1.3 (Alcatel, 1996, p4; Lockyear 1998, p53).

The system is composed of a three level control hierarchy:

- **Management Level** - System Management Centre (SMC) providing non-vital supervisory control for operator interface, automatic scheduling and regulation;
- **Operations Level** - Vehicle Control Centre (VCC) providing vital control of vehicle and track equipment;
- **Activation Level** - providing Vehicle On-Board Controllers (VOBCs), as well as interlocking and inductive loop transmissions.

(Alcatel 1996, p3; Lockyear 1998, pp53 & 55).

The communication system is used to perform 3 functions:

- Transmission of continuous digital data between the control centre and trains;
- Voice communication between the central control and passengers or operators on trains;
- Vehicle positioning by detection of loop crossovers every 25m.

(Kitchenside 1998, p193).

The continuous communication between VOBCs and VCCs includes details of line speeds, gradient profiles, maximum service brake rates, enable / close door commands, train depart commands, vehicle status and initiate / release emergency brake commands (Lockyear 1998, p55). The track to train data is transmitted at rates of 1200 bits per second on a 36kHz carrier frequency. Train to track data is transmitted at 600 bits per second on a 56kHz carrier frequency. In both cases, frequency modulation occurs at 0.4Hz.The high data exchange rates enable the system to be implemented with a high level of vehicle automation and control, including implementation as a moving block system (Alcatel 1996, p4).

Each VCC, operating on a GA 935 mini computer, can support up to 15 track loops (each of up to 3.2km in length) and 30 trains, monitoring the position, velocity and travel direction of each train (Alcatel 1996, p6-7; Lockyear 1998, p53). The VCCs contain 'guideway' data (station and point locations, loop positions, track gradients, line speeds etc.) to provide them with a model of the railway that they control. They are also programmed with relevant safety distances and service brake rates. In addition to this, they are interfaced to point and train detection equipment, and also receive vehicle position reports (Lockyear 1998, pp53-54). Based on this data, they are able to determine safe stopping points, maximum speeds and other commands that are then sent every second in a telegram to each train within their area of control (Kitchenside 1998, p193).

Each VOBC operates as a dual microprocessor unit, including redundant microprocessors that monitor each other's output commands to vehicle subsystems and messages to the VCC in order to check correspondence. If one microprocessor fails, control of the train is switched to a microprocessor in another vehicle of the train, or manual control can be exercised. VOBCs interpret commands from the VCC and control the vehicle's movements accordingly, transmitting vehicle position and speed, travel direction and subsystem status to the VCC (Alcatel 1996, p6). The trainborne system has no fixed geographical data of the railway but purely responds to instructions from the VCC, determining the vehicle's position by comparison of detected track loop transpositions, tacho-generator and accelerometer outputs (Lockyear 1998, p55).

Each vehicle can operate in either ATO or cab-signalling modes and, in both, the VOBC automatically applies the emergency brake if an overspeed condition is detected or if communication between a VCC and a train is lost (Alcatel 1996, p6; Lockyear 1998, p54). Following a loss of communication, the VCC prevents any other train's target point from advancing beyond the non-communicating vehicle's last reported position. On the Docklands Light Raliway system, overlaid axle counters also allow the system to track movement of an un-communicating train and continue to provide protection under a fixed block arrangement as it moves through the line under manual control (Lockyear 1998, p54).

When implemented on the Docklands Light Railway, this system initially suffered from poor reliability. Operational experience showed that the most significant threat to reliability was common mode software failures (Lockyear 1998, p60).

C7.8 VICTORIA LINE

The Victoria Line Automatic Train Control system contains both a safety system (ATP) and a train command system (ATO). The ATP element of the system is based on double rail coded track circuits, operating at 125Hz. The track circuit carrier signal is interrupted to produce pulses with equal 'on' and 'off' duration, at a rate of 120, 180, 270 or 420 pulses per minute (Challis 1976, p33; Smith 1966, p3). A separate code generator is used for each frequency, with the code to apply being selected according to the track circuit occupancy of the line ahead. Absence of a code causes automatic emergency brake application, making the system fail safe. (Challis 1976, p33).

The trainborne equipment detects the track circuit code by use of coils mounted immediately above the rails before the train's leading wheels (Smith 1966, p3). It then uses solid state circuits to control the train in accordance with signals received:

- The 120 code indicates that there is a train on the track circuit ahead. It is not recognised by the trainborne equipment and causes an immediate brake application when received;
- The 180 code will permit the train to travel at up to 22mph without motoring;
- The 270 code also has a limit of 22mph, but will permit motoring;
- The 420 code is used to permit the train to run at maximum speed.

(Challis 1976, p34; Kitchenside 1998, p191).

At reversing points, trains are required to operate in both directions over the same track circuit. In order to permit this, a feed and relay are connected to both ends of the track circuit. Both relays, but only one of the feeds, are used at any time, with the required feed being selected on the basis of the route set (Challis 1976, pp43-4).

In addition to the ATP codes, audio frequency generators are connected to one rail at appropriate locations, known as 'spots' (Kitchenside 1998, p191). The signals from 'spots' extend for about 10 feet and are detected by a second pair of coils at the front of the train. During an approach to a station, spot codes are applied to correspond with the ideal braking curve, at frequencies of 100Hz per 1mph of target speed. In addition, spot codes of 15kHz and 20kHz are used to activate coasting and to stop a train at a signal respectively (Challis 1976, pp 33, 66 & 70).

The ATO functionality controls the service brake on approach to station stops in accordance with received spot codes. If the train speed is found to be in excess of that represented by a spot code, the service brakes are applied harder. If the train is found to be going slower than the speed represented by a spot code, the service brakes are released (and, if supported by the track circuit code, traction power will be applied). If the train speed is found to be in excess of that represented by a track circuit code, the ATP functionality intervenes and an emergency brake application occurs (Smith 1966, p4).

In order to ensure safety, the ATP system operates using two overlaps per signal, calculated for different speeds. As with conventional trainstops, one overlap commences at the track circuit boundary by a signal or where a signal would be located if conventional signalling were being used. This is only of sufficient length to stop a train travelling at up to 25mph. If a signal is overrun whilst it is at danger, the train enters this overlap. It then recieves a speed code of 120 pulses per minute, which causes an automatic emergency brake application.

The second overlap commences at the point at which a train under normal approach conditions should be travelling at 22 mph. The track circuit between the signal and this point is fed with a 180 code whilst the signal is at danger, so that any train reaching this point at a speed above 22mph will be emergency braked. This second overlap is provided with sufficient length to ensure that a train entering the overlap at the maximum attainable speed will be stopped within it. This ensures that a train will not reach the reduced starter overlap at above 25mph (since the

180 code will not allow acceleration from the previously checked speed of 22mph) (Smith 1966, pp4-5).

In station areas, a more complex arrangement of overlaps is used. As with conventional signalling, a home signal is located a 25mph overlap distance in rear of the station. If a train approaches the home signal at red (because of a second train already in the platform) it will, therefore, be protected by the usual dual overlaps. However, in addition to this, the platform area is divided into several track sections. This means that if the station starter signal clears and the train in the platform begins to move off, a series of overlaps based around the multiple station track circuits can be activated. This has the effect of causing the ATP system to behave as if a conventional signalling system with three home signals were installed, even though only one home signal physically exists. The overlaps are located such that when the departing train leaves a track circuit, a 25mph overlap with a 120 track circuit code is set up behind it, whilst the track circuit before that is energised with a 270 code. This allows the following train to proceed past the home signal whilst continuing to enforce a safe separation between the two trains (Smith 1966, p6).

In case of failure of the ATO operation, a train can be driven manually at speeds of up to 22mph, as long as a valid track circuit code is still being received. If the track circuit code fails, the trains must be manually operated at a speed below 10mph. In order to support these fall back modes of manual operation, station starter, home and junction protection signals are still provided, along with other intermediate signals as required on long inter-station runs (Kitchenside 1998, p191). The signals have three aspects (red, white and green). Green has a conventional meaning (that the line is clear up to the end of the overlap of the next colour light signal), whilst a white aspect means that the line is clear up to the end of the overlap of the next 'ATP block'. To train drivers, a white aspect only gives authority for an automatically driven train to proceed. A manually driven train must wait for a green aspect (Challis 1976, p52).

C7.9 SRAWS SPEED SUPERVISORY SYSTEM

The basic SRAWS system included no speed supervision. However, other experiments were conducted in the Wilmslow area that included two trainborne computers, operating in a two out of two arrangement (Alston et al. 1971, p33). These were used to calculate speed profiles and distance to go information based on:

- Details of train weight and other characteristics fed into the computer by punched cards;
- Geographical information (including signal locations, speed restrictions and gradients) fed into the computer from a continuous track loop.

  (Kitchenside et al. 1998, pp205-6)

Separate track loops were used for each signal section. The SRAWS transmissions continued as before, with a second channel to transmit the additional data required, operating with frequency shift keying on a carrier frequency of 65kHz. By this means, the system was able to transmit details of the next four signal locations (including the average gradients on approach to each), the next four speed restrictions (including distance and average gradient to their commencement), the present speed limit and the direction to which this data applied (Alston 1971, p31-3).

In addition to the two SRAWS displays, a third 'supervisory speed' display was used, consisting of a standard speedometer with indicator lamps at 5mph intervals around it and two digital displays. The indicator lamps were illuminated to indicate the maximum permissible speed and the next target speed, whilst the digital displays indicated the distance to the next target and the current train location (Alston et al. 1971, p33).

Despite the major advances in protection offered by these trials, the system once again did not progress to general adoption, due to the limitations of the technology then available (Kitchenside et al. 1998, pp205).

C7.10 ETCS LEVEL 2

As with ETCS Level 1, the level 2 system can designed as an overlay on a conventional signalling system. The system uses conventional trackside equipment (track circuits, axle counters, interlocking, etc) to provide train detection and train separation functions (Thomas et al 1992, p28). Level 2 is based on continuous GSM-R radio coverage for track to train and train to track transmissions, supplemented by intermittent transmissions from Eurobalises (which are mainly used for location referencing) (Tomlinson 2001, p12; UNISIG 1999, chapter 2 p20). The radio system used is GSM-R, with allocated frequency ranges of 876 to 880 MHz (trackside to train) and 921 to 925 MHz (train to trackside). This allocation is sufficient to permit 20 GSM channels, each of which can support 7 simultaneous train connections (O'Neill 1999, p134).

The same trainborne equipment will be used for level 2 operation as for level 1 (with the obvious addition of radio communication equipment), allowing fitted trains to operate in areas with trackside fitment of either of the two systems. The trainborne system also continues to operate as it did under level 1 (providing supervision of speed and movement authority). However, under level 2 the information required to derive movement authorities is obtained from the central interlocking rather than the signal head, with the appropriate movement authority then being transmitted to the train by radio. This approach means that, when sufficient signal blocks are clear ahead of the train, the level 2 system need not be limited by the underlying signalling's normal speed and movement authority limits. It can instead provide the driver with an in-cab movement authority that extends beyond the range that can be displayed by signal aspects and, by so doing, authorise operation at higher speeds (where appropriate). The inclusion of features such as this marks a divergence between the functionality of level 1 and 2 systems. Since the underlying signalling can no longer provide all of the information required by the driver, the in-cab display of a level 2 system needs to indicate authoritatively the current movement authority and associated information (such as speed restrictions), acting as an in-cab signalling system (Bott 2000, p23; O'Neill 1999, p27).

In 2002, approximately 21% of the UK main line network was fitted with electronic interlockings capable of supporting Level 2 functionality. This figure was expected to increase to 68% by 2020 (Waboso 2002, p51).

Under a level 2 system, it is possible to suppress or remove lineside signals and operate purely on the basis of in-cab signalling (UNISIG1999, chapter 2 p20). In order to do this, all trains must be fitted before removal of the conventional track based signals (Waboso 2002, p7).

Since level 2 systems are based on radio transmission, information (including movement authorities) can be updated at any time or location (Woodland[2] 2000, p5). If signals are not suppressed, this capability should avoid most of the inconsistencies between conventional multiple aspect signal indications and in-cab display information that are to be expected in a level 1 system. Some discrepancies will still occur due to transmission delays, but the duration of these should be short (Bott 2000, p23).

Whether lineside signals are suppressed or not, the driver will still be expected to drive 'head-up' (observing the state of the line ahead) even though signalling information is available in the cab. However, since the in-cab display provides the most authoritative source of signalling information, this means that changes to the permitted speed or movement authority need to be drawn to the driver's attention. This will be achieved by use of an audible warning (Bott 2000, p23).

Studies for both NS Reizigers and Railtrack have predicted that Level 2 systems implemented as overlays onto lineside signalling systems will offer reduced capacity over conventional fixed block signalling (and even level 1 systems) due to the system's radio transmission delays (Cusk 2001, p6; Holtzer 1999, p589). However, when implemented as a cab-signalling system without lineside signals, Level 2 is expected to offer potential for capacity increase. Studies in the UK have suggested that an increase of 5% could be obtained (possibly up to 10% with complimentary investment to remove existing signalling constraints, change operating rules and amend advisory speed control), whilst the European Commission has predicted capacity enhancement of 25-30% for high speed and 12-16% for conventional rail lines (Scherp 2003, p5; Waboso 2002, p47).

The ETCS level 2 system is currently being tested across Europe, including at Old Dalby in the UK. The first deployment in the UK is currently scheduled for passenger operation on the Cambrian line in 2008 (National ERTMS Programme Team 2004, p24).

C7.11 ETCS LEVEL 3

As with Level 2, ETCS Level 3 is a GSM-R radio based train control system, supplemented by intermittent transmissions from Eurobalises (for location referencing). In a level 3 system, train location and mitigation of the risks involved in the loss of train integrity are performed by a trackside Radio Block Centre (RBC), in co-operation with the trains themselves (which send position and train integrity reports to the RBC). Each RBC controls the locking and release of routes as trains pass through its area. They do this by transmitting movement authorities and track descriptions to each train, based on the information received from all trains under their control (Tomlinson 2001, p13; UNISIG 1999, chapter 2 pp22-3).

Level 3 is intended for operation without lineside signals (UNISIG 1999, chapter 2 p22). In consequence of this, the in-cab display is the driver's primary source of signalling information (supplemented by lineside markers, where required).

These characteristics have a number of significant implications:

- Lineside equipment can be minimised (by reduction in lineside train detection equipment and removal of signals), offering reliability and maintainability advantages;
- The reported train location will be subject to odometery errors, which must be tolerated within the safety logic. In effect, this requires an allowance to be added to the train length, increasing headway time. Unless lineside train location devices (such as track circuits, position detectors and axle counters) are retained in junction areas, this could also result in delayed junction release;
- Unless fallback modes of operation are provided, all trains operating in level 3 areas must be fully equipped with functional ETCS trainborne equipment. Operational procedures must also be in place to deal with ETCS equipment failure on trains within a Level 3 area, or trains that arrive at a Level 3 area without operational equipment and have to be diverted elsewhere.

(Based on O'Neill 1999, pp33-35)

The same trainborne equipment will be used for level 3 operation as for levels 1 and 2 (with the obvious addition of train integrity proving equipment), allowing fitted trains to operate in areas fitted with the trackside equipment of any of the three systems.

Functionally, level 3 will be the same as level 2, with the trainborne system continuing to provide supervision of both speed and movement authority. However, since level 3 systems do not rely upon geographically fixed train detection systems or fixed lineside signals, alterations to the signalling (such as track layout changes) will become easier, in most cases only requiring data changes (O'Neill 1999, p35). It will also be possible to implement them as 'moving block' if required (see section 4.2.4) (Thomas et al. 1992, p28). Whether or not this is done, the driver will still be expected to drive 'head-up' (observing the state of the line ahead). As with level 2, all new information requiring the driver's attention therefore needs to be notified audibly (Bott 2000, p28).

Theoretically, the use of moving block train separation should make the capacities achievable under level 3 higher than those conventionally achieved by fixed block lineside signalling systems, speeds of operation being limited only by infrastructure and train capabilities rather than the designed signal separations. Moving block arrangements would also offers improved perturbation recovery (optimising capacity to actual running speeds, rather than designed line speeds), along with the ability to have bi-directional working on all lines (O'Neill 1999, pp34-35). However, studies for NS Reizigers have suggested this may not be true in all cases (Holtzer 1999, p589).

There are currently no definite plans for deployment of level 3 in the UK and "there is no direct work on level 3 in Europe at the present time" (Waboso 2002, p8).

C7.12 MAIL RAIL

The mail rail system is unique in the author's experience, in that it relies upon control of the traction supply to provide both automatic train protection and operation. From the outset (in 1927) mail rail has operated as a driverless system without signals. Originally, routes in station areas were controlled by mechanical levers, with the sections between stations operating as automatic sections. The levers could only be moved from normal to reverse if:

- There was no mechanical locking from conflicting routes;
- Track circuits in the section were clear;
- Points in the route were detected in the correct position.

Once these conditions had been satisfied, a feed operated a relay which then operated the contactors that applied traction current to the rails (Cotton 1994, p1).

In order to provide control of train movements, the track was divided into blocks, each having its own track circuit and conductor rail section. Occupation of a track circuit could therefore be used to provide electrical locking of route levers and also to isolate the section in rear's conductor rail (Dapre 1999, p13).

When a train completed a route, the lever could be moved partially back, but a lock prevented returning it to normal. A circuit then confirmed that the traction breakers had operated properly and the current had been removed from the starting track. On confirmation, the lever was released to return to the normal position (Cotton 1994, p1).

Depending on the number of track sections set for a train to use, either 440V DC or 150VDC was applied to each section. 440V was sufficient to power a train to a maximum speed of 35mph, whilst 110V would only enable movement at 7mph. In the absence of any traction supply, the train's brakes automatically applied (Kitchenside 1983, p198). Low speed shunting moves were enabled by a short application of 440V to start the train from rest, followed by 150V for the remainder of the route. (Cotton 1994, p1). The speed of approach could therefore be controlled and inherent automatic train protection provided without the need for any trainborne intelligence.

In 1993 This system was replaced with a dual processor based centralised control system (operating in hot standby) to replace the mechanical lever frames that had been located at each station. This utilises the same inputs and control mechanisms, but provides control by VDU displays and automatic route setting facilities (Cotton 1994, pp2-4).

C7.13 AATC

AATC was developed by Harmon Industries (later GE Transportation Systems), initially for application on the Bay Area Rapid Transit (BART) system in San Francisco. It is based on an enhanced position location reporting system (EPLRS) that was initially developed for the US Department of Defence, utilising 2.4 GHz spread spectrum radio transmissions (Anon[(2)] 1999, p17; Noffsinger 2001, p4).

AATC divides the track into zones, each controlled by a station computer and interface controller (interfacing to the station interlocking). The station computer is connected to two radio sets, and additional wayside radio sets are distributed throughout the area of control, such that two to four other units can receive every message transmitted by any radio unit. Each train that will operate under AATC is also fitted with two radio sets at opposite ends of the train (Anon[(2)] 1999, p17).

All AATC radio sets (both trackside and trainborne) are synchronised across the whole network and combine within any control zone to form a single TDMA communications network, communicating with a time division access (TDMA) protocol using 0.5s frames of 256 slots. Each radio unit has an assigned transmission slot within each frame, allowing up to 20 trains to be tracked and communicated with in any control zone (Anon[(4)] 2002, p2; Anon[(2)] 1999, pp17-18).

If a radio unit receives only one copy of a message, it forwards it to other radio units in its area. If it receives two or more copies, it only forwards them if they are identical. This forwarding action replaces the need for hard-wired connections between the units, allowing the system to be completely wireless. The use of assigned time slots also enables detection of any unit's failure by its neighbouring units (due to the absence of a transmission in its time slot), which is automatically reported. The communication chain then re-configures itself to bypass the faulty unit (Anon[1] 1999, p5; Anon[2] 1999, pp17-19).

The station computers each consist of two sets of three processor modules (one for network management services, one for vital station computer functions and the third for non-vital station computer functions). They track train movements, manage the radio network and calculate speed commands for transmission to the trains, as well as providing an interface to neighbouring station computers, to the control centre and to station ATO equipment. One set of computers acts as the master, whilst the other is in hot standby, with only the master unit outputting its results to the communication sub-system. The non-vital processor of each unit monitors the operational status of the other unit and reports any faults. In case of a failure, the standby unit immediately becomes the master (Anon[1] 1999, p3; Anon[2] 1999, pp17-18).

As a train enters a control zone, or a replacement wayside unit is installed, the network services processor assigns a reporting slot in the TDMA network to each unit. (Anon[2] 1999, p19).

Each trainborne radio connects to an interface controller, the two controllers being interconnected. Each interface controller then connects to an ATC unit which provides the trainborne ATP and ATO functionality (Anon[2] 1999, p17).

By listening to transmissions by the nearest 4 wayside radio units, trainborne radio units are able to extract message time stamps and compare these with the messages time of arrival to determine the distance between themselves and each wayside unit. These calculated distances are then transmitted to the trackside radio units every 0.5s. By this means, every train's front and rear locations can be determined by the station controllers at 0.5s intervals and to a guaranteed accuracy of +-15feet (4.6m). Since the station controllers know the exact location of radio units and the line topology, they are also able to determine the level of uncertainty contained within the calculated locations. This means that the locations can then be used to calculate and transmit new speed commands to all trains, with a 1mph (1.6km/h) resolution, based on braking curve calculations that include a variable safety margin allowance (Anon[1] 1999, pp1, 5; Anon[2] 1999, p17; Anon[4] 2002, p2; Appendix A interview with Mervyn Parvard). The system can thus operate with moving block train separation.

In addition to the main components of the AATC system, the BART system has been fitted with a broken rail detection system track circuits have been used to provide track locking in points areas (Noffsinger 2001, p3).

APPENDIX D - STATION STARTER SIGNALS

This appendix is a continuation of section 7.3.1.

In order to prove that consecutive trains stopping at a station will not find the signals controlling departure from the station restrictive to their progress, two elements must be considered:

1. The aspect that must be displayed by the starter when the following train is ready to depart if it is to be allowed to proceed unrestricted by the leading train up to and beyond the point where it attains line speed;
2. The most restrictive aspect that could be displayed by the station starter when the following train is ready to depart if the lead train has continued to travel normally along the line from the moment when the following train reached the first home signal sighting point and observed a green aspect.

D.1 Acceptable Proceed Aspects from a Station Starter

In order to determine the first of these elements, it is necessary to consider how far the lead train would be in front of the following train when the following train attains line speed for each possible starter aspect. The following train will be unrestricted by the lead train, showing that the aspect being considered is acceptable, if this distance is not less than the plain line train following headway distance.

In order for the starter to display a green aspect on an 'n' aspect signalling system, the lead train must have reached a point given by:

$$H_n = \frac{(n-1)V^2}{2(n-2)b} + O + L \qquad \text{(where n>2)} \qquad \text{Equation D1. 1}$$

This is equivalent to plain line headway distance ahead, without any allowance for sighting distance (since that is not required by a train starting from rest).

To display a more restrictive aspect with line speed V, the separation can be reduced such that:

$$Lead\ Train\ Start\ Location = \frac{(n-m)V^2}{2(n-2)b} + O + L \quad \text{(where n>2 and m< n)} \qquad \text{Equation D1. 2}$$

In this equation, the green aspect is displayed if 'm' is equal to 1. For each integer increase in the value of m above 1, the next most restrictive aspect is displayed (until m is one integer below n, which represents the most restrictive proceed aspect).

Since the lead train began from rest in the platform, its speed (V_{ftd}) at the time that the following train is ready to depart will be given by:

$$V_{ftd} = \begin{cases} \left\{V\right\} & if \ \sqrt{2a\left(\dfrac{(n-m)V^2}{2(n-2)b} + O + L\right)} \\ \left\{\sqrt{2a\left(\dfrac{(n-m)V^2}{2(n-2)b} + O + L\right)}\right\} & otherwise \end{cases}$$

Equation D1. 3

The following train will take a time given by V/2 to reach line speed, covering a distance given by $V^2/2a$. In this same time interval, the lead train will travel a distance:

$$Lead\,Train\,Dist = \begin{cases} \left\{\dfrac{V^2}{2a}\right\} & if \ V \le \sqrt{2a\left(\dfrac{(n-m)V^2}{2(n-2)b} + O + L\right)} \\ \left\{\dfrac{V^2 - 2a\left(\dfrac{(n-m)V^2}{2(n-2)b} + O + L\right)}{2a} + \dfrac{V\sqrt{2a\left(\dfrac{(n-m)V^2}{2(n-2)b} + O + L\right)}}{a}\right\} & otherwise \end{cases}$$

Equation D1. 4

The separation once both trains have achieved line speed will, therefore, be given by:

$$Line\ Speed\ Separation = \begin{cases} \left\{\dfrac{V^2}{2a} + \dfrac{(n-m)V^2}{2(n-2)b} + O + L\right\} & if \ V \le \sqrt{2a\left(\dfrac{(n-m)V^2}{2(n-2)b} + O + L\right)} \\ \left\{\dfrac{V\sqrt{2a\left(\dfrac{(n-m)V^2}{2(n-2)b} + O + L\right)}}{a}\right\} & \end{cases}$$

Equation D1. 5

The most restrictive aspect that can be used for a station start without causing restriction in the following train's movement further down the track can be found by comparison of these results with the train following headway distance in Equation 3-7. For realistic main line railway conditions (a = $0.4 m/s^2$, d = $0.5\ m/s^2$, s = 8s, O = 182m and L=200m, with line speeds of between 30 and 140mph), it is always acceptable to proceed on a yellow aspect. This is still the case if the equations are revised to assume that signal separations are increased by a third of the calculated braking distance, as permitted under Railtrack rules for signal location (Fleming 2000, p5).

D.2 Expected Aspect When Train Ready To Depart From a Single In-line Station

Returning to the second element, the most restrictive aspect that could be displayed by the station starter when the following train is ready to depart can be determined by considering the location and movement of the lead train. The aspect will be determined by the distance that the lead train travels between the moment that the following train reaches the first home signal sighting point (where the driver observes a green aspect) until the moment that the following train is ready to depart from the station. Under the worst case, the lead train will start from a location just clear of the station starter overlap (thus allowing the first home signal to have just cleared to a green aspect). The time taken by the following train to come to a stand at the station starter signal will be given by:

$$TimeToStop = SightingTime + TimeToBrake + TimeAtLineSpeed = S + \frac{V}{b} + \frac{V}{(n-2)2b} \qquad \text{Equation D2.1}$$

The time after which it will be ready to depart is, therefore, given by

$$TimeToDepart = S + \frac{(n-1.5)V}{(n-2)b} + Dwell \qquad \text{Equation D2.2}$$

The initial speed of the lead train will be:

$$V_{ftd} = \begin{matrix} \{V\} \ \ if \ \ V \leq \sqrt{2a(O+L)} \\ \left\{\sqrt{2a(O+L)}\right\} \ \ otherwise \end{matrix}$$

Equation D2.3

When the following train is ready to depart, it will therefore be located:

$$Lead\ Train\ End\ Location = \begin{cases} \left\{V\left(S+\frac{(n-1.5)V}{(n-2)b}+Dwell\right)\right\} & if\ \ V \leq \sqrt{2a(O+L)} \\ \left\{\begin{matrix} \sqrt{2a(O+L)}\left(S+\frac{(n-1.5)V}{(n-2)b}+Dwell\right)+ \\ \frac{a}{2}\left(S+\frac{(n-1.5)V}{(n-2)b}+Dwell\right)^2 \end{matrix}\right\} & if\ \ V > \left(\sqrt{2a(O+L)}+S+\frac{(n-1.5)V}{(n-2)b}+Dwell\right) \\ \left\{= SV+\frac{(n-1.5)V^2}{(n-2)b}+DwellV+\frac{V\sqrt{2(O+L)}}{\sqrt{a}}-O-L-\frac{V^2}{2a}\right\} & otherwise \end{cases}$$

Equation D2.4

The most restrictive aspect that could be displayed by the station starter signal when the following train is ready to depart can be found by comparison of these results with the train following headway distance in Equation 3-7. For realistic main line railway conditions (a = 0.4m/s^2, d = 0.5 m/s^2, s = 8s, O = 182m, L=200m and dwell = 60s, with line speeds of between 30 and 140mph), the most restrictive aspect displayed will always be green. This is still the case if the equations are revised to assume that signal separations are increased by a third of the calculated braking distance, as permitted under Railtrack rules for signal location (Fleming 2000, p5).

D.3 Summary of Results for Station Starter Signal With All Trains Stopping – Single In-line Station

Under the typical conditions considered in this appendix, the most restrictive aspect ever displayed to a train ready to depart a station with a single in-line platform is a green aspect. Under the same conditions, trains can always proceed from a 4-aspect station starter on a double yellow aspect and a 3-aspect station starter on a yellow aspect without experiencing any restriction due to the train in front. The station starter and subsequent signals are, therefore, not as significant for headway as the station home signal.

D.4 Expected Aspect When Train Ready To Depart From a Station with One In-line and One Off-line Platform per Direction

Where there is more than one platform from which the lead train can depart, the aspect to be expected on the station starter signal when a train becomes ready to depart will still be determined by the location and movement of the preceding train. However, it is no longer valid to assume that the worst case separation on approach to the platform will be when the lead train is located just clear of the station starter overlap as the following train arrives at the sighting point of the first home signal. In the worst case, the first home signal could display a clear aspect when the lead train is just route setting time clear of the station approach junction. In such a case, the lead train would not yet have completed its dwell in the station (and may not even have stopped in it yet) when the following train can receive a green aspect on the first home signal. As a result of this, the location of a preceding train may not be so far ahead of the platform when a following train becomes ready to depart as it would be in the case of a single in-line platform.

The time taken by the following train between arrival at a green home signal and its being ready to depart the station will still be as stated in Equation D2.2. However, the initial speed of the lead train is no longer so easy to determine. It will depend on the distance between the clearing point and the station stop, as well as the duration of the route-reset delay. The train may still be braking to rest in its platform, or may be part way through its station dwell. If, for the sake of example, it is assumed that the trains have a length of 200m, that the clearance point is 225m before the station stop point, that it takes 10 seconds for route reset behind the lead train and that the train decelerates at 0.5 m/s^2:

- The expected speed of the lead train on clearing the clearance point would be 5m/s;
- The train would come to rest at the station stop point 10 seconds later (just as the first home signal clears to a green aspect);

The lead train would then remain in the platform for the duration of its required dwell, before accelerating away for a time that can be derived from Equation D2.2 to be:

$$Lead\ Train\ Acceleration\ Time = S + \frac{(n-1.5)V}{(n-2)b} \qquad \text{Equation D4. 1}$$

At the end of which, the following train will be ready to depart the station. The location of the lead train at this time will be given by:

$$Lead\ Train\ End\ Location = \begin{cases} \left\{\frac{a}{2}\left(S+\frac{(n-1.5)V}{(n-2)b}\right)^2\right\} & if\ \ V \geq a\left(S+\frac{(n-1.5)V}{(n-2)b}\right) \\ \left\{\frac{V^2}{2a}+V\left(S+\frac{(n-1.5)V}{(n-2)b}-\frac{V}{a}\right)\right\} & if\ \ V \leq a\left(S+\frac{(n-1.5)V}{(n-2)b}\right) \end{cases}$$

Equation D4. 2

Where V is the line speed.

The most restrictive aspect that could be displayed by the station starter signal when the following train is ready to depart can be found by comparison of these results with the train following headway distance in Equation 3-7. For realistic main line railway conditions ($a = 0.4m/s^2$, $d = 0.5\ m/s^2$, s = 8s, O = 182m, L=200m and dwell = 60s, with line speeds of between 30 and 140mph), the most restrictive aspect displayed under 4-aspect signalling will be double yellow and under 3-aspect signalling, yellow. This is still the case if the equations are revised to assume that signal separations are increased by a third of the calculated braking distance, as permitted under Railtrack rules for signal location.

The example used in this section was carefully selected to simplify the mathematics involved in analysis. Varying the distance that the train must travel between the clearance point and station stop point would allow the first home signal to display a green aspect to a following train before the lead train had stopped in the platform (for an increase) or part way through its dwell (for a decrease). However, it would also increase (or decrease) the time required by the following train to come to a stand in the platform. For trains with equivalent braking performance, the difference in run time over the additional distance would be equal for both trains and varying this parameter would not affect the aspect displayed on the station starter when the following train completes its dwell time.

The other significant variable in the calculation is the route reset delay. If this were to be less than 10 seconds, the following train would be able to receive a green aspect at the first home signal before the lead train had stopped in the platform. If it were greater than 10 seconds, the lead train would be part way through its dwell time before the first home signal cleared to green. In either case, the variation would directly increase/decrease the lead train acceleration time.

With a route reset delay of 0 seconds, the most restrictive aspect displayed under 4-aspect signalling with minimum signal separation would be double yellow and under 3-aspect signalling, single yellow. However, if signal separations were increased by a third of the calculated braking distance, the most restrictive aspect displayed would be a single yellow under both 3 and 4-aspect signalling. This is true at some speeds for any route-reset delay below 8 seconds, as shown in Table D 1.

<table>
<tr><th rowspan="2">Route Reset Delay</th><th colspan="2">Most Restrictive 3-Aspect Starter Aspect</th><th colspan="2">Most Restrictive 4-Aspect Starter Aspect</th></tr>
<tr><th>Minimum Signal Separation</th><th>Maximum Signal Separation</th><th>Minimum Signal Separation</th><th>Maximum Signal Separation</th></tr>
<tr><td>8s</td><td>G 30<V<53mph
Y 53<V<140mph</td><td rowspan="9">Y 30<V<140mph</td><td>G 30<V<53mph
YY 53<V<140mph</td><td>YY 30<V<140mph</td></tr>
<tr><td>7s</td><td>G 30<V<44mph
Y 44<V<140mph</td><td>G 30<V<44mph
YY 44<V<140mph</td><td>YY 30<V<135mph
Y 135<V<140mph</td></tr>
<tr><td>6s</td><td>G 30<V<36mph
Y 36<V<140mph</td><td>G 30<V<36mph
YY 36<V<140mph</td><td>YY 30<V<108mph
Y 108<V<140mph</td></tr>
<tr><td>5s</td><td rowspan="6">Y 30<V<140mph</td><td rowspan="6">YY 30<V<140mph</td><td>YY 30<V<81mph
Y 81<V<140mph</td></tr>
<tr><td>4s</td><td>YY 30<V<54mph
Y 54<V<140mph</td></tr>
<tr><td>3s</td><td rowspan="4">Y 30<V<140mph</td></tr>
<tr><td>2s</td></tr>
<tr><td>1s</td></tr>
<tr><td>0s</td></tr>
</table>

Table D 1: Two Platform Expected Starter Aspects Under 3 and 4-Aspect Signalling

D.5 Expected Aspect When Train Ready To Depart From a Station with Three or more Platforms per Direction of Travel

With 3 platforms per direction of travel, the first train would be stopping in the platform as the second train arrived. By the time the third train arrived, the first train would be well through its dwell time – or even leaving the platform. The relationship between the first two trains and that between the 2nd and 3rd trains would still be the same as that of the one platform case. The provision of 3 or more platforms per direction would, therefore give the same expected starter signal aspects when trains are ready to depart as the case of a station with one in-line and one off-line platform per direction.

D.6 Summary of Results for Station Starter Signal With All Trains Stopping – multiple Platforms per Direction

Under the typical conditions considered in this appendix, it is possible for a train ready to depart from a platform at a station with multiple platforms per direction to receive an aspect that is more restrictive than would be required to maintain headway. The conditions under which this situation occurs can be seen in Table D 2.

<table>
<tr><th rowspan="2">Route Reset Delay</th><th colspan="2">Restrictive 3-Aspect Starter Aspect</th><th colspan="2">Restrictive 4-Aspect Starter Aspect</th></tr>
<tr><th>Minimum Signal Separation</th><th>Maximum Signal Separation</th><th>Minimum Signal Separation</th><th>Maximum Signal Separation</th></tr>
<tr><td>8s</td><td>G required, Y displayed 53<V<72mph</td><td rowspan="10">G required, Y displayed 30<V<72mph</td><td rowspan="10">Never too restrictive for headway</td><td>Never too restrictive for headway</td></tr>
<tr><td>7s</td><td>G required, Y displayed 44<V<72mph</td><td>YY required, Y displayed 135<V<140mph</td></tr>
<tr><td>6s</td><td>G required, Y displayed 36<V<72mph</td><td>YY required, Y displayed 108<V<140mph</td></tr>
<tr><td>5s</td><td rowspan="7">G required, Y displayed 30<V<72mph</td><td>YY required, Y displayed 81<V<140mph</td></tr>
<tr><td>4s</td><td>YY required, Y displayed 54<V<140mph</td></tr>
<tr><td>3s</td><td rowspan="4">YY required, Y displayed 30<V<140mph</td></tr>
<tr><td>2s</td></tr>
<tr><td>1s</td></tr>
<tr><td>0s</td></tr>
</table>

Table D 2: Multiple Platform Expected Starter Aspects Under 3 and 4-Aspect Signalling

Where this situation occurs, the station starter and subsequent signals are more significant for headway than the station home signal.

A typical route reset delay would be in the region of 8 seconds (TCS JPT 2002, Infrastructure work sheet). On this basis, and considering that the maximum permitted speed restriction through a passenger platform is 125mph, for a 4-aspect signalled station with multiple platforms per direction, the starter signal would be expected to be less restrictive to headway than the first home signal. However, in the case an equivalent station with 3-aspect signalling, it would be likely that the starter signal would be more restrictive to headway than the first home signal.

APPENDIX E - COMPARISON WITH ANALYTICAL RESULTS FROM PREVIOUS HEADWAY RESEARCH

E.1 Anon. 'Moving Block: Is It Worth the Money', 1967

The article by 'a leading signal engineer' quotes train following headways for 3, 4 and 'n' aspect signalling systems, along with moving block. Diagrammatic representations of plain line signalling sections are provided to explain the origins of these equations :

$$Ht_n = \frac{S + \left(\frac{n-1}{n-2}\right)B + O + L}{V} \qquad \text{(Anon 1967, p606)} \qquad \text{Equation E1. 1}$$

Since with equi-spaced signals in an 'n' aspect signalling system the braking distance, B, is equal to the signal separation, d, multiplied by (n-2), these can be seen to be equivalent to the equations derived for train following headway in chapters 3 and 4 (equations 3-11and 4-1). The anonymous author notes that under moving block 'O' becomes a safety margin, rather than an overlap, defining this as 5 seconds (which would equate to 279m at a speed of 125mph). No reasoning is given for the selection of this value, which seems to be rather large at high speeds. However, the explanation may lie in the fact that throughout the paper the author assumes a maximum line speed of 90mph, giving a safety margin of 200m, which seems far more reasonable (Anon. 1967, p606).

The author appears to assume that sighting requirements 'S' will be the same for fixed and moving block systems. In practice, there will be a subtle difference, since moving block implies in-cab signalling rather than lineside displays. As discussed in section 4.2.4, the reliance of moving block signalling on trainborne position measurement and reporting systems will also result in different delays and accuracy errors that must be considered when determining the sighting distance to assume in headway calculations.

The author does not develop the quoted headway equations to derive point headways for trains stopping at stations, passing through junctions or deviating from line speed in any other way.

However, textual descriptions and graphs of the effect of such deviations are given (Anon. 1967, p608). These appear to reflect the equivalent results derived in chapter 7.

E.2 Bergmann 'Generalized Expressions for the Minimum Time Interval between Consecutive Arrivals at an Idealized Railway Station', 1972

Bergmann takes the base assumption that successive vehicles on a railway must be separated by a gap that is not less than the following vehicle's instantaneous stopping distance. To this end, he considers brake build up delays and variation in train speeds, such that the minimum separation between trains must be given by the integral:

$$H = \int_{t}^{t+bbu} V(t)dt + Bd + L$$ (Bergmann 1972, p329) Equation E2. 1

where L refers to the leading vehicle, V, Bd and bbu to the following vehicle.

Bergmann notes that an integral cannot be easily evaluated and, since the speed may well vary during the brake build up, makes the assumption that bbu = 0 when calculating headway distance. He then adds the actual value of bbu on to the eventually calculated headway time.

As with the author's own modeling, Bergmann assumes that all trains have the same performance and characteristics and apply acceleration and braking at constant rates following the brake build up delay. However, he also assumes an idealised railway that makes no allowance for the technology used or the system delays that might be introduced as a result of that technology. Factors such as overlaps / safety margins, sighting time and system delays are therefore ignored, with the interesting exception of the allowance for brake build up delays already mentioned (Bergmann 1972, pp329-331).

Within his paper, Bergmann quotes the equations for Headway time as being:

$$Ht = \frac{V}{2b_e} + \frac{V}{2b_o} + \frac{V}{2a} + \frac{L}{V} + dwell + bbu, \ for V < \left[\sqrt{\frac{2ab_e b_o L}{b_e b_o + ab_e - ab_o}} \right] and\, b_0 \le be$$

(Bergmann 1972, p333) Equation E2. 2

$$Ht = \frac{V}{b_o} + dwell + bbu + \sqrt{\frac{2}{a}\left(L + \frac{V^2}{2b_e} - \frac{V^2}{2b_0} \right)}, \ for V > \left[\sqrt{\frac{2ab_e b_o L}{b_e b_o + ab_e - ab_o}} \right] and\, b_0 \le be$$

(Bergmann 1972, p333) Equation E2. 3

where b_o is the operational brake rate and b_e the emergency brake rate

If the same rate is assumed for operational and emergency braking, or the 'safe' separation between trains is calculated on the basis of operational rather than emergency braking rates, this gives the equations:

$$Ht = \frac{V}{b} + \frac{V}{2a} + \frac{L}{V} + dwell + bbu \ \ for V, \left[\sqrt{2aL}\right]$$ (Bergmann 1972, p335) Equation E2. 4

$$Ht = \frac{V}{b} + dwell + bbu + \sqrt{\frac{2L}{a}}, \ for V < \sqrt{\left[\frac{2ab^2 L}{0}\right]}$$ (Bergmann 1972, p335) Equation E2. 5

These can be seen to be the same as the two conditions given in equation {7-21} in the author's thesis, with the addition of the allowance for bbu, but neglecting any allowance for overlap or sighting times.

Bergmann goes on to compare his equations with those of other preceding authors' papers on similar subjects. This comparison shows that:

- Equation E2.4 was also quoted, without derivation and neglecting the applicable limits for V, in a paper by McGlumphy entitled 'Optimising headway', presented to the American Transit Association group conference in May 1963 (Bergmann 1972, p336);
- Lang and Soberman published a book with the MIT Press in 1964 entitled 'Urban Rail Transit: Its Economics and Technology'. Their equations were based on a different assumption for minimum train separation: namely that the leading end of the following train and trailing end of the leading train could be allowed to momentarily touch as they respectively enter and depart the station area. This obviously gives a lower headway time of:

  $$H_t = 2\sqrt{L/a} + dwell$$ (Bergmann 1972, p337) Equation E2. 6

  where 'a' represents a common rate of acceleration and deceleration and the momentary 'touch' would therefore occur half way along the platform.

  This is quite obviously an oversimplification, but an interesting concept that represents even closer operation of trains than would be permitted under relative braking operation (since the vehicles are allowed to touch at speed).

E.3 Pearson, 'Moving Block Railways Signalling', 1973

When considering speed restrictions, Pearson assumes that the signal spacing (and therefore the minimum train following headway distance) will remain as a constant distance, sufficient for that required by operation at the full line speed. He then compares the distance over which a train must travel below line speed as a result of the speed restriction with the minimum train following headway distance, calling the difference 'W'. Where the restriction has a length R_L and the train following headway at line speed is Hn, W is then stated to be given by:

$$W = Hn - R_L - \frac{(V^2 - V_R^2)\times(a+b)}{2ab}$$ (Pearson 1973, p14) Equation E3. 7

If W is positive, the first train will regain line speed and travel a distance W before a following train reaches the location at which it must begin braking for the speed restriction. The headway for the speed restriction is, therefore, given to be:

$$Hn_R = \frac{W}{V} + \frac{R_L}{V_R} + \frac{(V - V_R)\times(a+b)}{ab}$$ (Pearson 1973, p15) Equation E3. 8

If W is negative, the effect of the restriction extends beyond the line speed train following headway distance. It then becomes necessary to consider whether or not a train will clear the restriction before a following train arrives at the point where it must commence braking for the restriction. Pearson defines a value 'W_2' to consider this, where:

$$W_2 = Hn - R_L - \frac{(V^2 - V_R^2)}{2b}$$ (Pearson 1973, p15) Equation E3. 9

If W2 is positive, the first train will have passed through the speed restriction by a distance equal to W_2 and will be accelerating back up to line speed when a following train reaches the critical location at which it must begin braking for the speed restriction. Pearson, therefore, quotes the speed of the train at this instant as being:

$$V_{INST} = \sqrt{V^2 + (2aW)} \qquad \text{(Pearson 1973, p15)} \qquad \text{Equation E3. 10}$$

Pearson notes that W is negative at this point (which would make the equation invalid). In actual fact, it is the modulus of W that should be used here, since it is not actually W that defines the distance remaining between the train and the point at which line speed would be retained. It is the difference between Hn and the distance that would be required between the point at which trains must start braking for the restriction and the point at which they regain line speed (in this case, -W).

On this basis, the headway for the speed restriction where W is negative and W2 is positive is given by Pearson as:

$$Hn_R = \frac{R_L}{V_R} + \frac{(V - V_R)}{b} + \frac{V_{INST} - V_R}{a} \qquad \text{(Pearson 1973, p15)} \qquad \text{Equation E3. 11}$$

If W2 is negative, the first train would still be within the speed restriction, a distance '-W_2' from the end of it, when the following train reached the location at which it must begin braking for the speed restriction. The headway would then be given by:

$$Hn_R = \frac{R_L + W_2}{V_R} + \frac{(V - V_R)}{b} \qquad \text{(Pearson 1973, p15)} \qquad \text{Equation E3. 12}$$

The same equation is also quoted for moving block signalling, albeit with a different value for W_2 (Pearson 1973, p83).

Unfortunately, all of these equations assume that a train is required to remain at the restricted speed only for the length of the speed restriction. In practice, the train would need to remain at the restricted speed until the whole train had cleared the restriction. That is, for an additional distance equal to the train length. Therefore, they all predict a lower minimum headway time than would actually apply.

The assumption that the required headway distance will remain the same when a speed restriction is encountered as it would with a constant maximum speed is also open to question. Since the Railtrack standards require fixed block signal spacing to be determined on the basis of the permissible speed, the signal separation would be expected to reduce in response to a long permanent speed restriction. This reduction would only be expected to commence for signals located within the speed restriction. Braking distance from line speed (or the previous higher speed restriction) would need to be maintained for the route ahead of the signal before the restriction. The headway calculations outlined by Pearson are, therefore, correct for the entry to a permanent speed restriction, or for the duration of a temporary speed restriction, but not for the duration of an extended permanent speed restriction. They are also unable to account for the potential effect of multiple speed restriction changes in close proximity.

Pearson's approach to nomenclature cannot be intuitively followed. He also significantly constrains the scope of his analysed scenarios, in order to simplify the derivation of equations and subsequent analysis. Due to the difficulty of deciphering his equations, this is not entirely obvious to a casual reader. The result of his approach is an analysis of limited application, which he none-the-less portrays as being representative of all scenarios. The author of this thesis was not convinced of the validity of this assumption.

E.4 Pope R, 'The Institution of Railway Signal Engineers Booklet No. 27: Signalling the layout British Railways Practice', 1975

Pope quotes equations for 3 and 4-aspect signalling in accordance with BR practice, quoting the headway time as being:

$$Ht_3 = \frac{S + 2B + O + L}{V} \qquad \text{(Pope 1975, p19)} \qquad \text{Equation E4.1}$$

$$Ht_4 = \frac{S + 1.5B + O + L}{V} \qquad \text{(Pope 1975, p19)} \qquad \text{Equation E4.2}$$

If it is considered that where signals are located optimally in accordance with the braking distance, B is equal to d in 3-aspect signalling and to 2d in 4 aspect signalling, these can be seen to be equivalent to those derived for train following headway in chapters 3 and 4 (equations 3-11and 4-1). Pope does not develop these equations to derive point headways for trains stopping at stations, passing through junctions or deviating from line speed in any other way. However, textual descriptions of the effect of such deviations are given (Pope 1975, pp20-53).

E.5 Nock O S, 'Railway Signalling', 1980

Nock also quotes equations for train following headways:

$$Ht_{2(RapidTransit)} = \frac{S + X + O + L}{V} \qquad \text{(Nock 1980, p7)} \qquad \text{Equation E5. 1}$$

A sighting distance in the region of B+ 30% is assumed for rapid transit 2-aspect signalling.

$$Ht_{2(MainLine)} = S_t + \frac{B + X + O + L}{V} \qquad \text{(Nock 1980, p7)} \qquad \text{Equation E5. 2}$$

$$Ht_3 = S_t + \frac{2B + O + L}{V} \qquad \text{(Nock 1980, p9)} \qquad \text{Equation E5. 3}$$

$$Ht_4 = S_t + \frac{1.5B + O + L}{V} \qquad \text{(Nock 1980, p9)} \qquad \text{Equation E5. 4}$$

A sighting time of 10 seconds is assumed for BR main line 2, 3 and 4-aspect signalling

These equations are once again equivalent to those derived for main line train following headway in chapters 3 (equations 3-1 and 3-11).

Nock also considers the situation where a train operating at lower speeds than the design speed of a 4-aspect signalling section can treat the double yellow aspect as if it were a clear aspect, quoting the train following headway in this situation as:

$$Ht_{4(3)} = S_t + \frac{B + O + L}{V_{act}} \qquad \text{(Nock 1980, p11)} \qquad \text{Equation E5. 5}$$

This is equivalent to the equation derived in section 7.2.1 (equation 7-6).

Nock does not develop these equations to derive point headways for trains for trains deviating from line speed. However, textual and diagrammatic descriptions of the effect of such deviations are given for in-line station stops (Nock 1980, pp20-27). Textual descriptions of the effects on headway of passing through junctions, provision of loop platforms at stations and terminal stations are also provided (Nock 1980, pp28-34 & 42-49).

E.6 Gill, 'Computer-Assisted Design of Optimised Signalling Layouts for Rapid Transit Railways', 1986

In his thesis, Mr Gill quotes an equation for the fixed block headway time of an 'n'-aspect fixed block signalling system (neglecting equipment response delays) as:

$$H_n = \frac{(n-1)}{(n-2)}\left(\frac{V_{\max}}{2b}\right) + \frac{S+O+L}{V_{\max}} \qquad \text{(Gill 1986, p4-11)} \qquad \text{Equation E6. 1}$$

This equation assumes operation at the maximum line speed at all times, but could easily be adapted to represent operation at a lower speed. It is also identical to equation 4-1 (if the braking distance is expanded to $V^2/2b$) that was derived in chapter 4.

Equations for headway under various forms of moving block are also discussed. The first form considered is moving space block (see Appendix F). Neglecting equipment response delays and jerk limiting, Gill quotes the headway distance of this arrangement to be:

$$H_{msb} = \left(\frac{V_{\max}^2}{2b}\right) \qquad \text{(Gill 1986, p4-24)} \qquad \text{Equation E6. 2}$$

The train length is ignored in the equation, as are the need for a sighting time and safety margin.

Gill then considers 'pure' moving block, quoting the headway distance (neglecting equipment response delays) to be:

$$H_m = \left(\frac{V_{act}^2}{2b}\right) \qquad \text{(Gill 1986, p4-26)} \qquad \text{Equation E6. 3}$$

The train length, sighting time and safety margin are still neglected in this equation. Allowing for these omissions, this equation can be seen to be equivalent to those derived in section 4.2.4.

Gill then considers moving time block (see Appendix F). Neglecting equipment response delays, he quotes the minimum separation between trains operating under this arrangement to equal:

$$H_{mtb} = \left(\frac{V_{\max} \times V_{act}}{2b}\right) \qquad \text{(Gill 1986, p4-26)} \qquad \text{Equation E6. 4}$$

Once again, the train length, sighting time and safety margin are neglected and, unfortunately, Gill does not even attempt to postulate what the benefit of such an implementation might be.

The discussion of headway within this thesis appears to be more a reflection of material read during Mr Gill's literature review than original work. The equation quoted for fixed block 'n' aspect headways is correct (although no allowance is made for system delays), but is stated rather than derived. It is also an equation that has appeared in earlier works. The equations for moving block arrangements are very week, ignoring all system delays and margins (and the accompanying text does not inspire confidence that Mr Gill actually understood the arrangements that he was describing). The thesis, therefore, offers little significant information to the reader.

E.7 Gill et al., 'Computer-based Optimisation Techniques For Mass Transit Railway Signalling Design', 1992

In their paper on 'computer based optimisation techniques for mass transit railway design', Messrs Gill and Goodman developed equations for headway and train frequency, with specific reference to rapid transit railways. Their analysis began by considering the theoretical maximum headway for a railway during plain line train following operation:

$$Ht = \frac{B+L}{V} = \frac{V}{2b} + \frac{L}{V}$$ (Gill et al. 1992, p263) Equation E7. 1

This is the same as the basic equations for moving block headway derived in chapter 4 (equation 4-2), if the sighting time and overlap are both set to zero.

A similar equation is then quoted from a paper on 'Generalised Expressions for the minimum time interval between consecutive arrivals at an idealised railway station' (Bergmann 1972, pp327-341) for the case of point headway with all trains stopping:

$$Ht_{(Stopping)} = \frac{V}{b} + t_r + Dwell + \sqrt{\frac{2L}{a}}$$ Equation E7. 2

where t_r is the ATO system response delay (equivalent to a driver's sighting time)

This equation, which is used as the basis for the rest of the paper, is again the same as the basic equation for moving block headway already derived (equation 7-21), if the sighting time and overlap are both set to zero. The condition for low line speeds (when $V < \sqrt{2aL}$), however, is ignored and the fact that the quoted equation is only valid for $V \geq \sqrt{2aL}$ is not noted.

As put forward in the paper, the equation for stopping headway assumes that trains accelerate at a steady rate until they have travelled the distance required to clear the platform. However, for a low line speed, the trains may actually reach the line speed before clearing the platform and then travel at the line speed for the rest of the distance required. As this would take the train a longer time, adopting the proposed equation would provide too low an estimate of the headway. For the typical train length of 140m and acceleration rate of 1m/s^2 assumed within Messrs Gill and Goodman's paper, this makes their results inaccurate for line speeds below 60km/h. If their equations were adapted to include an overlap, the criteria for validity would become $V < \sqrt{2a(O+L)}$. Even a nominal overlap allowance of 80m would then make the results inaccurate for line speeds below 76km/h. This is clearly an error, since the maximum line speeds that can usually be expected on mass transit railways (the stated subject of the paper) only reach up to around 80km/h (see section 7.4).

The equation should more properly account for the time required by the train to reach line speed and then clear the remainder of the overlap at that constant speed, such that:

$$Ht_{(Stopping)} = \frac{V}{b} + t_r + Dwell + \frac{V}{2a} + \frac{O+L}{V}$$ Equation E7. 3

The authors of the paper move on from theoretical minimum (moving block) headways to consider multiple aspect systems with ATP. The analysis assumes that such systems are equi-block systems with an overlap formed by an additional track section. On this basis, the train following headway of an 'n' aspect system is defined as being given by:

$$Ht_n = \frac{(n+1)V}{2b(n-1)} + \frac{L}{V} \qquad \text{(Gill et al. 1992, p265)} \qquad \text{Equation E7. 4}$$

Even with the sighting time set to zero, this is different from the Equation 7-57 in chapter 7.

The difference between the equation quoted within the paper and that developed in section 7.5 can best be understood by considering the example of Figure E6-1, which is defined in the paper as representing a 5-aspect system. In the opinion of the author of this thesis, an error has been made in this definition. There are in fact 6 aspects in the example:

1. Danger (0/0);
2. Next signal at danger (V_3/0);
3. Next but one signal at danger (V_2/V_3);
4. Next but two signals at danger (V_1/V_2);
5. Next but three signals at danger (V_m/V_1);
6. All clear (V_m/V_m);

The paper also fails to note that the quoted equation would not be applicable for n equal to 1.

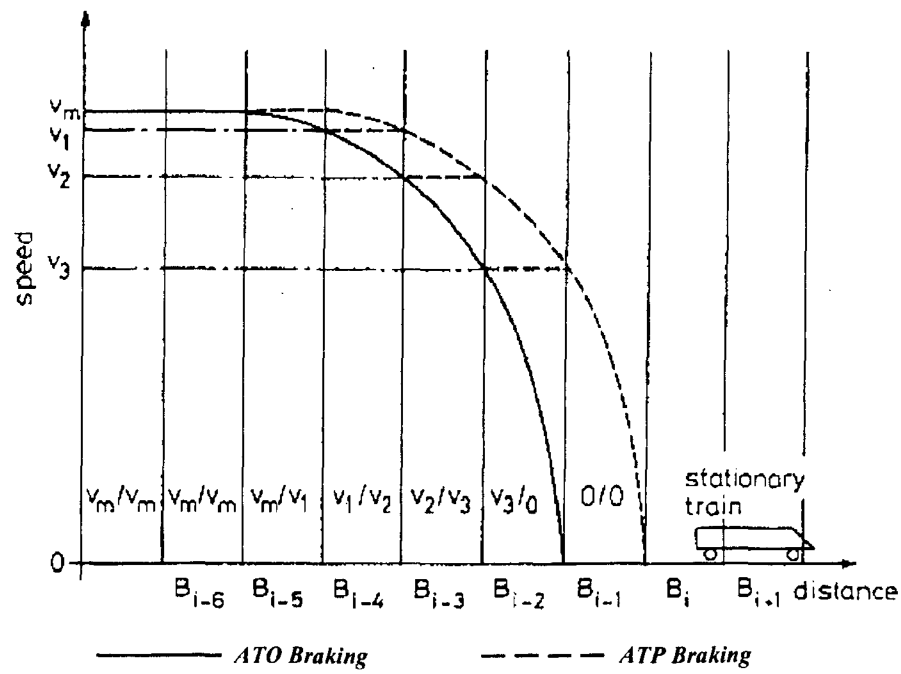


Figure E6-1: Track Circuit Codes and Relationships between ATO & ATP (Gill et al. 1992, p265)

Despite the errors discussed within this section, the paper offers an excellent outline of the factors to consider in analysis of a real railway's headway.

E.8 Hiroto T, 'Dynamic Response of Moving Block Signalling Systems', 1997

In his thesis, Mr Hiroto quotes an equation for the fixed block headway distance of a 4-aspect ATO based system (neglecting equipment response delays) as:

$$H_{4aspect} = \frac{5}{3}\left(\frac{V_{\max}^2}{2b}\right) \quad \text{(Hiroto 1997, p2-2)} \qquad \text{Equation E8. 1}$$

This equation is based on an equi-block approach, with the blocks each one third of the braking distance in length. This represents the type of signalling system used by systems such as TVM and the Central Line signalling, and previously described by fellow Birmingham University student Mr Gill and their supervisor Dr Goodman in their 1992 paper that has already been reviewed. In developing this equation Hiroto has made the same mistake as Gill and Goodman, describing a system that clearly has 5 aspects (which are all described by the author) as a 4-aspect system (p2-2 to 2-3). Hiroto has also over simplified the equation, with no allowance being made for train length.

Equations for headway under various forms of moving block are also discussed. The first form considered is moving space block (see Appendix F). Neglecting equipment response delays, Hiroto quotes the headway distance of this arrangement to be:

$$H_{msb} = \left(\frac{V_{\max}^2}{2b}\right) + SafetyM\arg in \quad \text{(Hiroto 1997, p2-5)} \qquad \text{Equation E8. 2}$$

The train length is again ignored in the equation, although is mentioned in the text accompanying it. There is also no mention of the need for a sighting time, but that could be accounted for as an equipment response time in an ATO arrangement and would therefore be covered by the exclusion comment.

Further analysis of Hiroto's statements about moving space block can be found in Appendix F.

The author then considers moving time block (see Appendix F). Neglecting equipment response delays, Hiroto quotes the minimum separation between trains operating under this arrangement to equal:

$$H_{mtb} = \left(\frac{V_{\max} \times V_{act}}{2b}\right) + SafetyM\arg in \quad \text{(Hiroto 1997, p2-6)} \qquad \text{Equation E8. 3}$$

Once again, the train length is neglected in this equation – and it does not ensure a constant time separation between trains. The headway time is, in fact, given by:

$$Ht_{mtb} = \left(\frac{V_{\max}}{2b}\right) + \frac{L + SafetyM\arg in}{V_{act}} \qquad \text{Equation E8. 4}$$

Unfortunately, Hiroto does not even attempt to postulate what the benefit of such an implementation might be.

Hiroto then considers 'pure' moving block, quoting the headway distance (neglecting equipment response delays) to be:

$$H_m = \left(\frac{V_{act}^2}{2b} \right) + Safety\,Margin \qquad \text{(Hiroto 1997, p2-7)} \qquad \text{Equation E8. 5}$$

The train length is still neglected in this equation. Allowing for this omission, and assuming that no allowance has been made for sighting the change in movement authority as that amounts to a system delay, this equation can be seen to be equivalent to those derived in section 4.2.4.

More detailed equations for pure moving block headway distance are quoted later in the Hiroto's thesis, during discussion of the effect produced by transmission delays. The minimum distance between the nose of a train and its limit of movement authority is then quoted as being:

$$Minimum\,Distance\,to\,LMA_m = (V(t) \times t_r) + \left(\frac{V(t)^2}{2b} \right) + \left(\frac{V(t) \times b}{2J} \right) + Safety\,Margin$$

(Hiroto 1997, p3-9) Equation E8. 6

Where: T_r is the transmission time delay (including equipment response delays);

b is based on a constant brake rate through the braking curve, modified by gradients and rotational inertia;

J is the jerk limit.

A further equation for the point headway a train stopping in a station under pure moving block is also derived in an appendix to the Hiroto's thesis. The headway time is then quoted as being:

$$H_{t\,m\,(Stopping)} = \frac{V}{b} + Dwell + \begin{cases} \sqrt{\dfrac{2(O+L)}{a}} & \text{For } V \geq \sqrt{2a(O+L)} \\ \dfrac{V}{2a} + \dfrac{O+L}{V} & \text{For } V \leq \sqrt{2a(O+L)} \end{cases}$$

(Hiroto 1997, p4-6) Equation E8. 7

This equation is the same as that derived in section 7.3.1 (equation 7-21), except that it neglects any allowance for sighting time.

Of perhaps most interest in this paper is the final moving block arrangement considered, that of relative moving block. This arrangement is based on the idea that a train cannot stop instantaneously so, if the speed and braking performance of the lead train are known, a movement authority can be given to the following train based on the earliest stopping point of the lead train, rather than its last reported location. Neglecting equipment response delays, Hiroto quotes the minimum separation between trains using this arrangement to be:

$$H_{rm} = \left(\frac{V_{act(following)}^2 - V_{act(lead)}^2}{2b} \right) + Safety\,Margin \qquad \text{(Hiroto 1997, p2-8)} \qquad \text{Equation E8. 8}$$

This equation applies so long as the lead train is travelling at a lower speed than the following train. If their speeds are the same, or the lead train is travelling faster, the minimum separation is equal to the Safety Margin (the train length is still neglected). Hiroto concludes that this arrangement (although that instinctively adopted by car drivers) would be unsafe if used for passenger transportation, but could be considered as satisfactory for freight traffic (pp2-8). Clearly, adopting this approach to train separation would offer possibilities for improving capacity. Further consideration has been given to this idea in Appendix F.

Whilst this thesis is by far the most comprehensive consideration of fixed and moving block headway analysis that has been reviewed, it has three major shortcomings:

- The first is that all system delays are neglected. Since the delays applicable to fixed and moving block operation are not necessarily the same, this means that a true comparison has not been achieved;
- The second is that the analysis only considers moving block and continuous fixed block ATP (actually stated to be ATO). No indication of the difference in impact that could be expected for intermittent ATP or pure conventional lineside signalling without ATP is given;
- The third (and, perhaps, most significant) is that the analysis ignored the impact of junctions, always looking at plain line sections. Since a junction could, in effect, be seen as a fixed block constraint to any moving block system, this could significantly reduce the advantage offered by moving block in any real system.

E.9 Holgate, 'Realising the Full Potential of Transmission Based Signalling', 1998

Holgate begins by considering moving block operation for non-stopping trains. Assuming that a safety margin 'O' must be maintained between trains at all times, he quotes the headway distance as being:

$$H_m = S_t V + B + O + L \qquad \text{(Holgate 1998, p3)} \qquad \text{Equation E9. 1}$$

The equivalent equation for an 'n' aspect system with an overlap of length 'O' is also quoted as being:

$$H_n = S_t V + \frac{(n-1)B}{(n-2)} + O + L \qquad \text{(Holgate 1998, p6)} \qquad \text{Equation E9. 2}$$

If it is considered that the sighting time S_t is equivalent to a sighting distance of S/V and that Headway distance can be converted into headway time by dividing by the speed, V, these equations are the same as the basic equations for 'n' aspect headway derived in chapter 4 (equation 4-1). Holgate does not expand on these equations to consider the case of stopping trains. However, he does use them to produce capacity / speed graphs that show:

- The relationship between capacity, speed and brake rate (producing a graph similar to Figure 5-13);
- The relationship between capacity, design speed and number of signalling aspects (producing a graph similar to that for train following headway in Figure 2-9);
- The relationship between capacity, actual speed and number of signalling aspects (producing a graph similar to Figure 7-1).

  (Holgate 1998, p9).

E.10 Uebel, 'Durchsatz von Strecken und Stationen', 1998

This paper, written in German, considers line and station capacity for operation under fixed block and moving block, with both absolute and relative braking distance separations.

Uebel quotes the headway time of absolute moving block to be given by:

$$H_{tm} = \frac{V_{act}}{2b} + \frac{O+L+Q}{V_{act}} + T \qquad \text{(Uebel 1998, p6)} \qquad \text{Equation E10. 1}$$

Where: Q is the location coarseness and T is the system's reaction time (including sighting time of system with manual driving).

Direct comparison of Mr Uebel's paper is difficult. However, assuming that 'Q' is equivalent to the distance travelled in the interval between determining location updates (Tlu) and that 'T' is equivalent to the sighting time, trackside and trainborne processing times and communication delays (that is, S + Pt + Ut), this equation can be seen to be a development of equation 7-10 in chapter 7, incorporating the system response delays considered in chapter 9.

Mr Uebel assumes that the value of T for fixed block operation is 10s. This is lower than the equivalent assumption of 14s made in chapter 9.

The equivalent equation for moving block operation with relative braking distance separations is also quoted as being given by:

$$H_{tm} = \frac{O+L+Q}{V_{act}} + T \qquad \text{(Uebel 1998, p6)} \qquad \text{Equation E10. 2}$$

Where: Vact is the speed of the following train';

In this case, Q can be taken as the equivalent to the train location update interval and error (Tlu + Tlu_e) and T as the equivalent to the warning margins, trackside and trainborne processing times and communication delays (that is, Tw + Tp+ Tip+ Pt + Tlu_t + Ut) in chapter 9. Mr Uebel assumes a value of 10s for T, significantly lower than the equivalent assumption of 20s made in chapter 9.

This equation clearly assumes that the lead train is travelling at a speed equal to that of the following train, such that there is no braking distance component in the required train separations. If the lead train were to be travelling at a lower speed, the difference in braking distance to rest would need to be accounted for. Similarly, if the lead train were travelling at a higher speed than the following train, the separation may not need to be even this high at the interval being considered as is implied by this equation. In all three cases, differences in gradients could also impact the relative braking distances of the two trains, requiring some amendment to the headway calculations. The equation quoted is, therefore, an over simplification of the problem (as can be seen by comparing equation E9.2 with equations 9-5 and 9-8 in section 9.6.

The author then moves on to quote an equation for headway time under moving block operation with absolute braking distance separation and a station stop:

$$H_{t\,m\,(Stopping)} = \frac{V_{2act}}{b_2} + Dwell + T + \begin{cases} \sqrt{\dfrac{2(O+L+Q)}{a_1}} & \text{For } V_{2act} \geq \sqrt{2\,b_t\,(O+L+Q)} \\ \dfrac{V_{2act}}{2a_1} + \dfrac{O+L+Q}{V_{2act}} & \text{For } V_{2act} \leq \sqrt{2\,b_t\,(O+L+Q)} \end{cases}$$

(Uebel 1998, p9) Equation E10. 3

The equation is equivalent to equation 7-21 in chapter 7, with the addition of the system response delays considered in chapter 9.

Mr Uebel assumes that the value of T for relative braking distance separations would be twice that for absolute braking separation (due to the need for communication with two trains). He therefore adopts the value of 20s, which is the same as that assumed in chapter 9.

Uebel notes that his comparison of moving block and relative braking headway shows that relative braking offers a gain in station headway, increasing line capacity but also increasing the risk of accidents. As a result, Uebel concludes that introducing relative braking operation should be possible on pure goods lines, but not for passenger traffic. He further states that the full extent of potential capacity improvements would only be achieved by running homogeneous, well regulated traffic, since unequal maximum speeds or braking characteristics significantly reduce the capacity achieved (Uebel 1998, p10).

Uebel also states that the critical factor in determining headway under moving block is the train's brake rate, but that the system reaction and communication delays become critical once relative braking separation is introduced (Uebel 1998, p10).

E.11 Mott D et al, 'A Goodness Function for Terminal Layouts of Urban Railways', 2003 (Draft)

Mott et al. consider the capacity effects of different terminal station arrangements. The calculation for train arrival times is given in Appendix 'A' as:

$$t = \frac{x_{DE}}{V} + \frac{V}{b}$$

(Mott et al. 2003, pp21-22) Equation E11. 1

Where X_{DE} represents the distance between the braking point to stop from line speed (V) before the platform entry junction and the braking point to stop from line speed in the platform.

Whilst it is noted in the text that the time taken to approach the station stop will depend on the turnout speed, this is not allowed for in the quoted equation (Mott et al. 2003, p21). There is also no allowance for system reaction delays in the point control and signalling systems used.

The equivalent equation for departure times is given in Appendix 'B' as:

$$t = \frac{V_R}{a} + \frac{x_{FB}}{V_R}$$

(Mott et al. 2003, p22) Equation E11. 2

Where X_{FB} represents the distance between the point at which a train accelerating from rest in the platform would attain the junction turnout speed (V_R) and the point at which it would clear the junction.

Here the impact of any turnout speed is allowed for. However, the equation still represents a simplification in that it assumes that the turnout speed will always be reached before the train clears the junction. In practice, this will not always be the case, especially for departure over a straight route (with no restriction below line speed) where the junction is located close to the platform. There is also again no allowance for system reaction delays in the point control and signalling systems used.

No equation is quoted for the overall terminal platform headway time. However, it is noted in the text that this will be represented by the sum of the train movement times and a layover time: including components for door opening; detraining of incoming passengers; boarding of outgoing passengers; door closing; time for the driver to vacate his cab, walk to the other end of the train and open up the other cab; time for the train's control system to reverse direction; any additional waiting time in the layover (Mott et al. 2003, p7).

General similarities can be seen between the approach adopted in this paper and that outlined in section 9.5.3. However the equations are greatly simplified.

APPENDIX F – ATP SYSTEM CAPACITY (FULL DERIVATIONS)

F1 INTRODUCTION

The process of developing the practical fixed block UK main line signalling headway equations (derived in chapter 8) to represent the headway of operation with ATP overlay and in-cab signalling systems was commenced in chapter 9. Due to the repetitive nature of deriving equations for multiple systems and scenarios, only plain line headways were considered in that chapter. Whilst this provided a useful outline of the approach adopted by the author in deriving equations for all ATP types and scenarios, the reader may find the full working and explanation of differences between each ATP arrangement for the remaining scenarios useful in evaluating the author's work. The remaining derivations are, therefore, included within this appendix.

F2 FIXED BLOCK ATP TRAIN CONTROL SYSTEM CAPACITY

F2.1 DIVERGING JUNCTION HEADWAY

The headway for a diverging junction under conventional signalling has already been considered in chapter 8 (section 8.3). Due to the presence of junctions, trains travelling over a junction may follow different paths, making the headway impact of ATP on junction scenarios more complex than on plain line scenarios.

F2.1.1 DIVERGING TRAIN FOLLOWED BY STRAIGHT THROUGH TRAIN - INTERMITTENT ATP OVERLAY

F2.1.1.1 NO APPROACH RELEASE OR SPLITTING DISTANT SIGNALS

Where the underlying signalling is not approach released, the ATP supervision curves for a straight through train that is following a diverging train will target a stopping point at the last junction protection signal, until the route has reset behind the lead train. The supervision criteria can then be updated to target the junction speed restriction and a subsequent stop at the first signal after the junction. This is represented in Figure F 2-1.

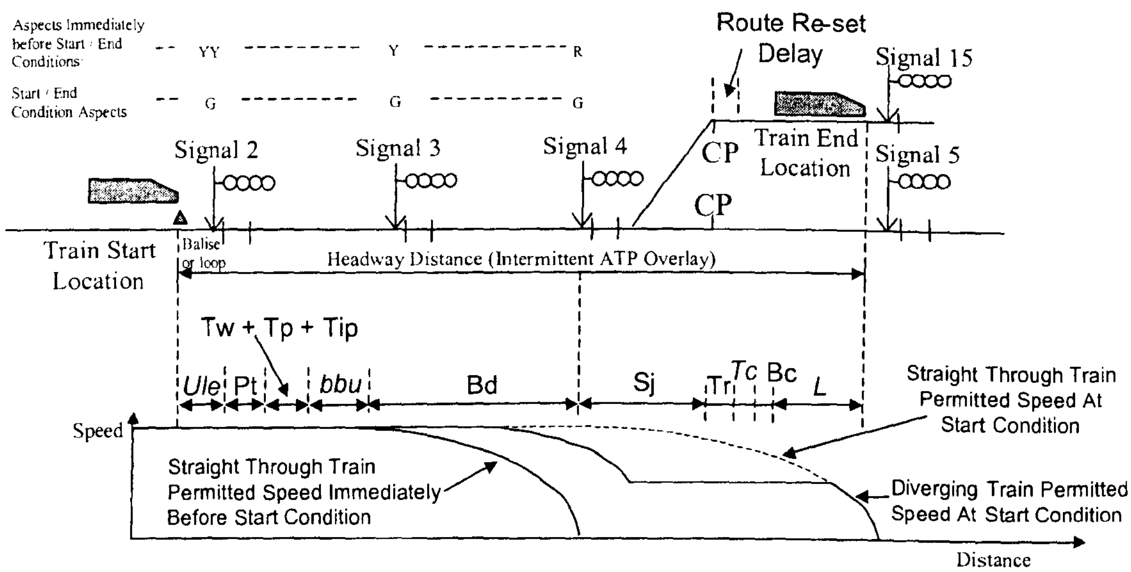


Figure F 2-1: Intermittent ATP Overlay Headway for Diverging Train Followed by Straight Through Train, No Approach Release

The headway time for this scenario is, therefore, given by:

$$H_{IATP-O} = \sum_{x_{start}}^{(x_{start}+Ule+Pt+Tw+Tp+Tip+bbu+Bd+Sj+Tr+Tc+Bc+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation F 2-1}$$

As in the case of plain line headway, if the ATP 'sighting' point (shown as the train start location in Figure F 2-1) is at or before the sighting point of the underlying signalling, the ATP will impose an increase in headway distance (which will in turn result in an increase in headway time). This may also be the case where the sighting point of the underlying signalling is slightly before the ATP sighting point (due to the balise/loop update delay). However, if the ATP sighting point is only required to be well within the sighting point of the underlying signalling, the ATP may not act as a constraint on achievable headway. In that case, the headway will be determined by applying equation 8-12 rather than Equation F 2-1.

The forward speed profile used to calculate $V_{act}(x)$ must again represent the speed profile that the driver of the leading (diverging) train would have to follow in order to comply with the lineside signal aspects, pre-empt intervention by the ATP system and comply with professional driving practice. In the case of a diverging junction with no approach release, the only restrictions to the trains' progress whilst they remain at headway separation would be the speed restrictions. As with plain line headway, this means that the driver's brake curve (at 90% of the service braking rate) will always be the most restrictive speed profile.

As discussed in section 8.3.10, when the approach to the junction is controlled by splitting distant signals, there is no approach control and the headway for a straight through train following a diverging train will still given by Equation F2.1. In this case, $V_{act}(x)$ will again be based on the driver's brake curve on approach to the junction speed restriction.

F2.1.1.2 APPROACH RELEASE FROM YELLOW, RED OR FLASHING YELLOW

The behaviour of the underlying signalling system when approach release from yellow, red or flashing yellow is introduced has already been considered in sections 8.3.2 to 8.34. It was shown that the critical headway where the first train diverges and the following train goes straight through is on approach to the junction, rather than through it. The scenarios to be considered are shown in Figure F 2-2, with the headway times still given by Equation F 2-1. If the headway permitted by the underlying signalling is more restrictive than the ATP headway, the headway time will still be given by Equation 8-12.

The forward speed profile used to calculate $V_{act}(x)$ must again represent the speed profile that the driver of the leading (diverging) train would have to follow in order to pre-empt intervention by the ATP system and comply with both the lineside signal aspects and professional driving practice. This would vary with the scenarios, as indicated by the speed / distance curves in Figure F 2-2.

Whilst the driver's brake curve on approach to a speed restriction has been taken to be 90% of the service-braking rate, it was noted in section 8.7 that, in accordance with professional driving practice, a braking curve on approach to a restrictive aspect should be based on 5%g. When approaching a signal that will be released from yellow, this means that $V_{act}(x)$ must be calculated on the basis of a 5%g braking curve to stop before the red signal ahead of it. This applies until the clearance point of the approach-released signal has been reached and the ATP supervision criteria known on-board the train have changed in response to the signal aspect release that this allows. The braking curve will then become 90% of service brake rate if the train is still travelling above the junction restriction speed. If the train speed falls below the value of the junction restriction whilst the driver is following the 5%g curve, the train may accelerate back up to junction speed once the signal has cleared.

If a signal that has been displaying a red aspect clears to a less restrictive aspect, an ATP system will continue to enforce the braking curve required for the red aspect until an update of supervision criteria occurs. In the case of an intermittent ATP system, such an update will only occur once the next balise or loop location is reached. In the intervening period, the train's speed may be forced well below that actually required by the underlying signalling system. In order to minimise this impact, many intermittent ATP systems include functionality that supports operation at a release speed on approach to a signal at red. Once activated, this functionality results in the trainborne ATP system supervising against the release speed, rather than the braking curve to stop at the signal. Release speeds may be implemented in a number of ways:

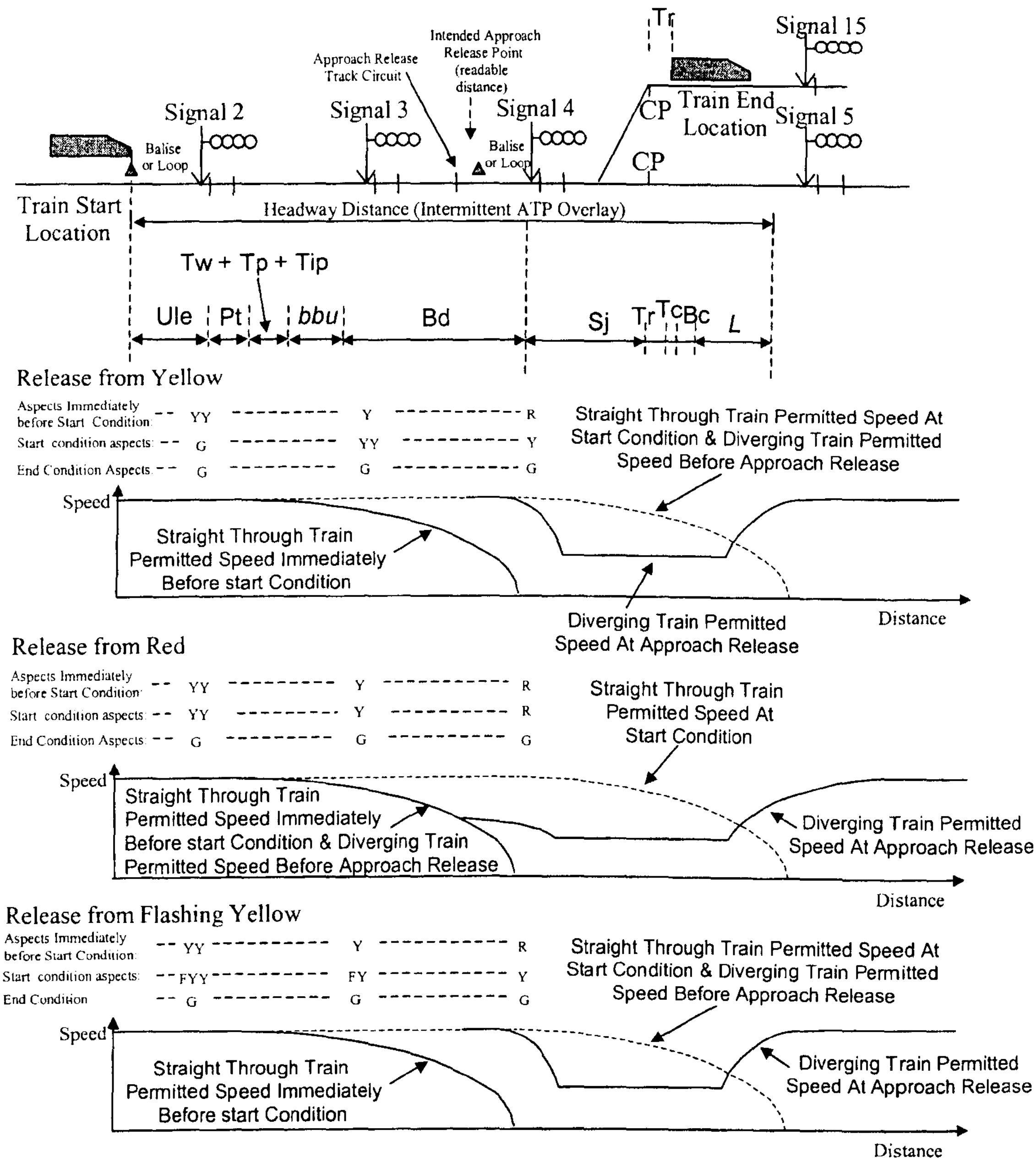


Figure F 2-2: Intermittent ATP Overlay Headway for Diverging Train Followed by Straight Through Train, with Approach Release

1. No release speed (the train must follow the braking curve required to ensure stopping at or before the red signal);

2. A fixed release speed for all signals. Supervision against the braking curve is released whenever the train falls below this speed or once the train reaches the location at which the braking curve intersects with the release speed;
3. A calculated release speed for each signal, based on the signal's overlap length and local gradients, such that the train with the poorest braking performance accepted for the line can be brought to rest within the overlap by a brake application that is triggered on passing the signal balise/loop whilst it is still indicating a red aspect. The value of the release speed must be transmitted to the train in advance of it being required. Supervision against braking curves can then be released once the train reaches the location at which the braking curve intersects with the release speed;
4. A release speed to be calculated on-board, based on the signal's overlap length and local gradients, as in the case of a calculated release speed for each signal. This approach permits allowance for the specific train performance characteristics, allowing higher release speeds for trains with better braking performance. The overlap length and gradients must be transmitted to the train in advance of the release speed being required, in order to allow for calculation. Supervision against braking curves can then again be released once the train reaches the location at which the braking curve intersects with the release speed.

When approaching a signal that will be released from red, $V_{act}(x)$ must be calculated at each location on the basis of the most restrictive profile out of:

- A 5%g braking curve to stop before the red signal, applied only until the approach-released signal has been released and either the ATP supervision criteria known on-board the train have changed in response to this or release speed supervision has been activated;
- A 90% of service brake rate curve to the junction speed.

Once again, if the train speed falls below the value of the junction restriction whilst the driver is following the 5%g curve, the train may accelerate back up to junction speed once the signal has cleared. However, if the train is above the junction speed but below the 90% service brake curve, the train should coast until further braking is required.

In the case of approach release from flashing yellow, the flashing aspects on approach to the junction advise the driver that the train will be taking the high-speed turnout ahead. However, prior to release of the fixed yellow aspect on the junction signal, the driver is still unable to tell whether the signal after the junction is actually displaying a red aspect. This means that $V_{act}(x)$ must be calculated at each location on the same basis used for approach release from yellow. Since approach release from flashing yellow is only used where the conditions required for approach release from yellow can not be met, it is more likely that the braking curve to junction speed will be significant at some point during the approach.

F2.1.2 DIVERGING TRAIN FOLLOWED BY STRAIGHT THROUGH TRAIN - CONTINUOUS ATP OVERLAY

F2.1.2.1 NO APPROACH RELEASE OR SPLITTING DISTANT SIGNALS

In this scenario, supervision curves for a straight through train that is following a diverging train across a junction with no approach control on the underlying signalling will target a stopping point at the last junction protection signal, until the route has reset behind the lead train. The supervision criteria can then be updated to target the junction speed restriction and a subsequent stop at the first signal after the junction. This is represented in Figure F 2-3.

The headway time for this scenario is, therefore, given by:

$$H_{CATP-O} = \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+Sj+Tr+Tc+Ut+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation F 2-2}$$

If the ATP 'sighting' point (shown as the train start location in Figure F 2-3) is at or before the sighting point of the underlying signalling, the ATP will impose an increase in headway distance which will in turn result in an increase in headway time. This may also be the case where the sighting point of the underlying signalling is slightly before the ATP sighting point, due to the balise/loop update delay. However, if the ATP sighting point is only required to be well within the sighting point of the underlying signalling, the ATP may not act as a constraint on achievable headway. In that case, the headway will be determined by applying equation 8-12 rather than Equation F 2-2.

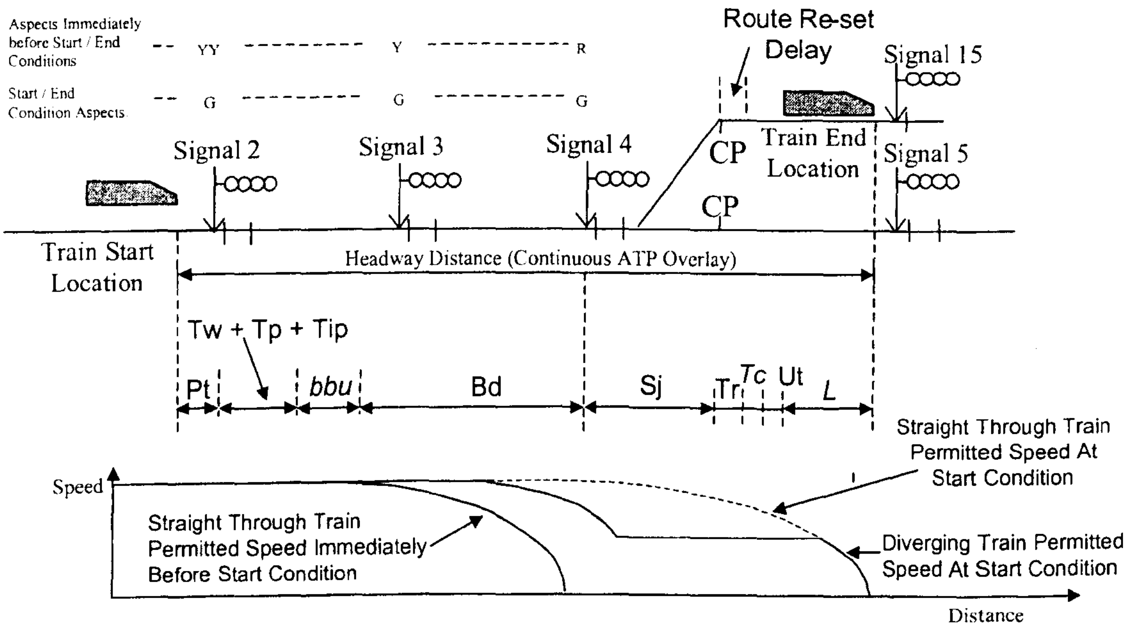


Figure F 2-3: Continuous ATP Overlay Headway for Diverging train Followed by Straight Through Train, No Approach Release

As in the case of an intermittent overlay ATP system, forward speed profile used to calculate $V_{act}(x)$ must represent the speed profile that the driver of the leading (diverging) train would have to follow in order to comply with the lineside signal aspects, pre-empt intervention by the ATP system and comply with professional driving practice. Once again, where no approach release is applied, this would be determined based on the driver's brake curve of 90% of the service brake rate on approach to the junction speed restriction.

As discussed in section 8.3.10, when the approach to the junction is controlled by splitting distant signals, there is no approach control and the headway time for a straight through train following a diverging train with still be given by Equation Equation F 2-2. In this case, $V_{act}(x)$ will again be based on the driver's brake curve on approach to the junction speed restriction.

F2.1.2.2 APPROACH RELEASE FROM YELLOW, RED OR FLASHING YELLOW

The behaviour of the underlying signalling system with approach release from yellow, red or flashing yellow has already been considered in sections 8.3.2 to 8.3.4. It was shown that the critical headway where the first train diverges and the following train goes straight through is on approach to the junction, rather than through it. The scenario to be considered is shown in Figure F 2-4.

The headway time for all of these scenarios is still given by Equation F 2-2. If the headway permitted by the underlying signalling is more restrictive than the ATP headway, it is given by Equation 8-12.

The forward speed profile used to calculate $V_{act}(x)$ must again represent the speed profile that the driver of the leading (diverging) train would have to follow in order to pre-empt intervention by the ATP system and comply with both the lineside signal aspects and professional driving practice. This would vary with the scenarios, as indicated by the speed / distance curves in Figure F 2-2 and subsequent textual description in section F2.1.1.2, since the supervision curve desired would not change between intermittent and continuous update.

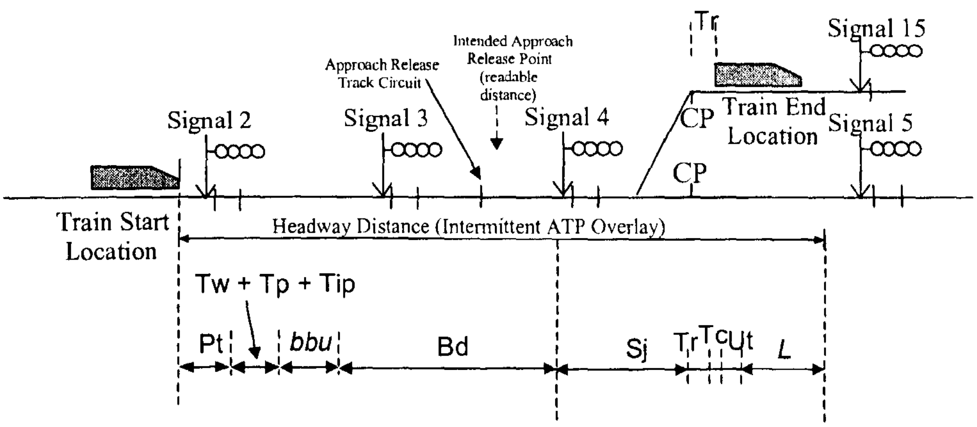


Figure F 2-4: Continuous ATP Overlay Headway for Diverging Followed by Straight Through Trains with Approach Release

F2.1.3 DIVERGING TRAIN FOLLOWED BY STRAIGHT THROUGH TRAIN - CONTINUOUS IN-CAB ATP

Since lineside signals are not required with continuous in-cab displays, the headway distance is always as shown in Figure F 2-5 and there is no need to consider the effects of approach control.

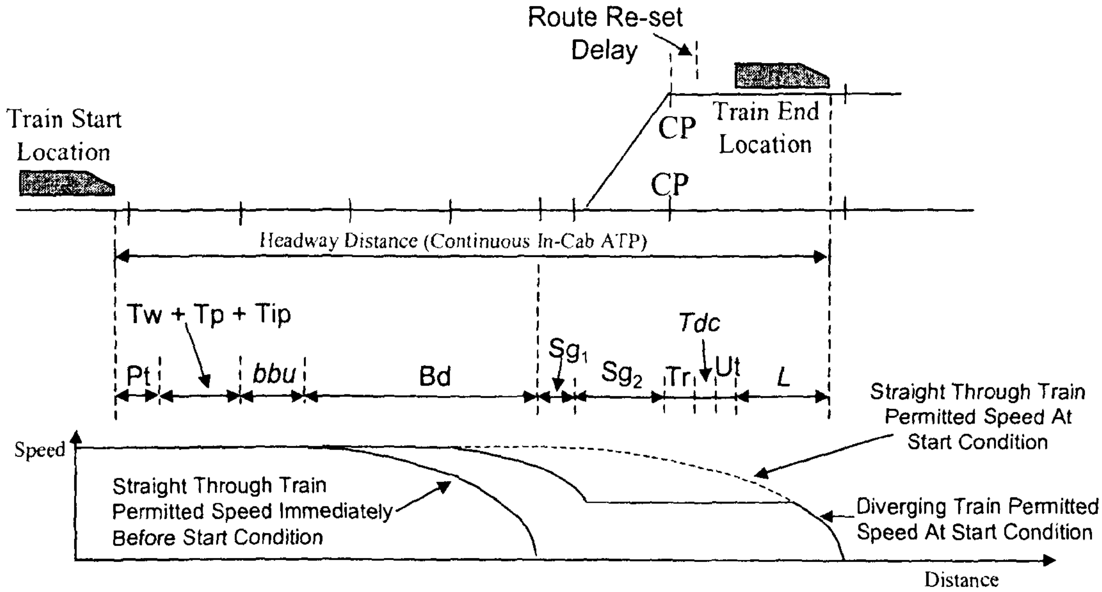


Figure F 2-5: In-Cab ATP Headway for Diverging Train Followed by Straight Through Train

It should be noted that in Figure F 2-5 the section Sg_2 must be cleared by the leading (diverging) train before the route can be reset behind it for a following movement in the straight through direction. The movement authority of the train behind cannot be advanced beyond the start of Sg_1 until the reset has occurred, as there is no overlap available beyond the Sg_1/ Sg2 boundary in the intended direction of travel. In effect, this means that so long as all track sections are at least the equivalent of an overlap length, the diverging followed by straight through headway of a continuous in-cab ATP system differs from that of a continuous ATP overlay where no approach control is applied only by the difference (if any) between their track section lengths and the delays that they experience between clearance of the junction track section and updated supervision criteria becoming available to the Trainborne systems. The headway time is, therefore, given by:

$$H_{CATP-In} = \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+Sg_1+Sg_2+Tr+Tdc+Ut+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation F 2-3}$$

It should be noted that this is actually a longer time than the equivalent plain line headway, by Tr (the junction reset time), since the track section Sg_2 is not available to be included in the following train's movement authority until the route has reset.

F2.1.4 CONSECUTIVE DIVERGING TRAINS - INTERMITENT ATP OVERLAY

F2.1.4.1 NO APPROACH RELEASE OR SPLITTING DISTANT SIGNALS

In section 8.3.6, it was noted that the headway of consecutive trains taking the diverging route at a junction with no approach release is the same as that for a plain line section with an equivalent speed profile. This is also the case for a junction controlled by splitting distant signals, which imposes no approach control. In both scenarios, the headway time for a straight through train following a diverging train can, therefore, be calculated using equation 9-1.

F2.1.4.2 APPROACH RELEASE FROM YELLOW, RED OR FLASHING DOUBLE YELLOW

As discussed in section 8.3.7, with the introduction of approach control, the critical underlying signalling headway distance for consecutively diverging trains travelling through a junction commences at the planned clearance point of the junction signal. Since the approach released signal cannot clear for a train operating at headway until it reaches this point, the train will be forced to follow the braking curve imposed by the controlled aspect on approach to it. Once the planned clearance point is reached, the junction signal will be released to display whatever aspect is actually permitted by the state of the line ahead.

The trainborne supervision equipment of an intermittent ATP overlay system will be unaware of the change in signal aspect that occurs at the planned clearance point and will therefore continue to enforce the braking curve appropriate for the controlled aspect until it receives and processes updated supervision criteria from a balise or loop. This means that the effective sighting point of the ATP system will be the ideal balise/loop location, a distance equivalent to the balise/loop message update delay after the underlying signalling system's planned clearance point. The headway distance will then extend a distance equivalent to the balise/loop message update delay after the end of the conventional signalling headway distance. This scenario is represented in Figure F 2-6.

The train will be forced to follow the ATP system's restrictive supervision curve (based on the controlled aspect) until the updated supervision criteria have been received and processed on-board. If the balise /loop location is ideal, this means that the train's speed will be restricted by the ATP system for the duration of the combined balise update and trainborne processing delays (Bc+Pt) beyond the underlying signalling system's planned clearance point. If the actual balise /loop location is not ideal, the restrictive supervision will extend for a further distance equivalent to the balise/loop location error (Ule).

It should be noted from Figure F 2-6 that, due to the variation in operating speed through the line, the distances travelled by the lead and following train during the balise/loop update time (shown as Bc_1 and Bc_2) may differ. Therefore, the headway time under intermittent ATP overlay operation is given by:

$$Ht_{IATP-O} = \sum_{x_{start}}^{\left(x_{start}+(S_A-Bc_1)+O+Tc+Bc_2+L+\sum_{p=4}^{p=n+3} Sg_p\right)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation F 2-4}$$

Where x_{start} is the ideal balise/loop location (Bc_1 beyond the intended clearance point of the underlying signalling junction signal).

The forward speed profile used to calculate $V_{act}(x)$ must again represent the speed profile that the driver would have to follow in order to pre-empt intervention by the ATP system and comply with both the lineside signal aspects and professional driving practice. This would vary with the scenarios, as indicated by the speed / distance curves shown in Figure F 2-6 and the discussion on professional driving practice in section F2.1.1.

Whilst the distance covered by the trains may differ between Bc_1 and Bc_2, the times determining those distances are the same. If the three distances within Equation F 2-4 (namely Bc_1, Bc_2 and Tc) are removed from the summation and instead added as their natural time elements, the two Bc values cancel and the equation becomes identical to Equation 8-14.

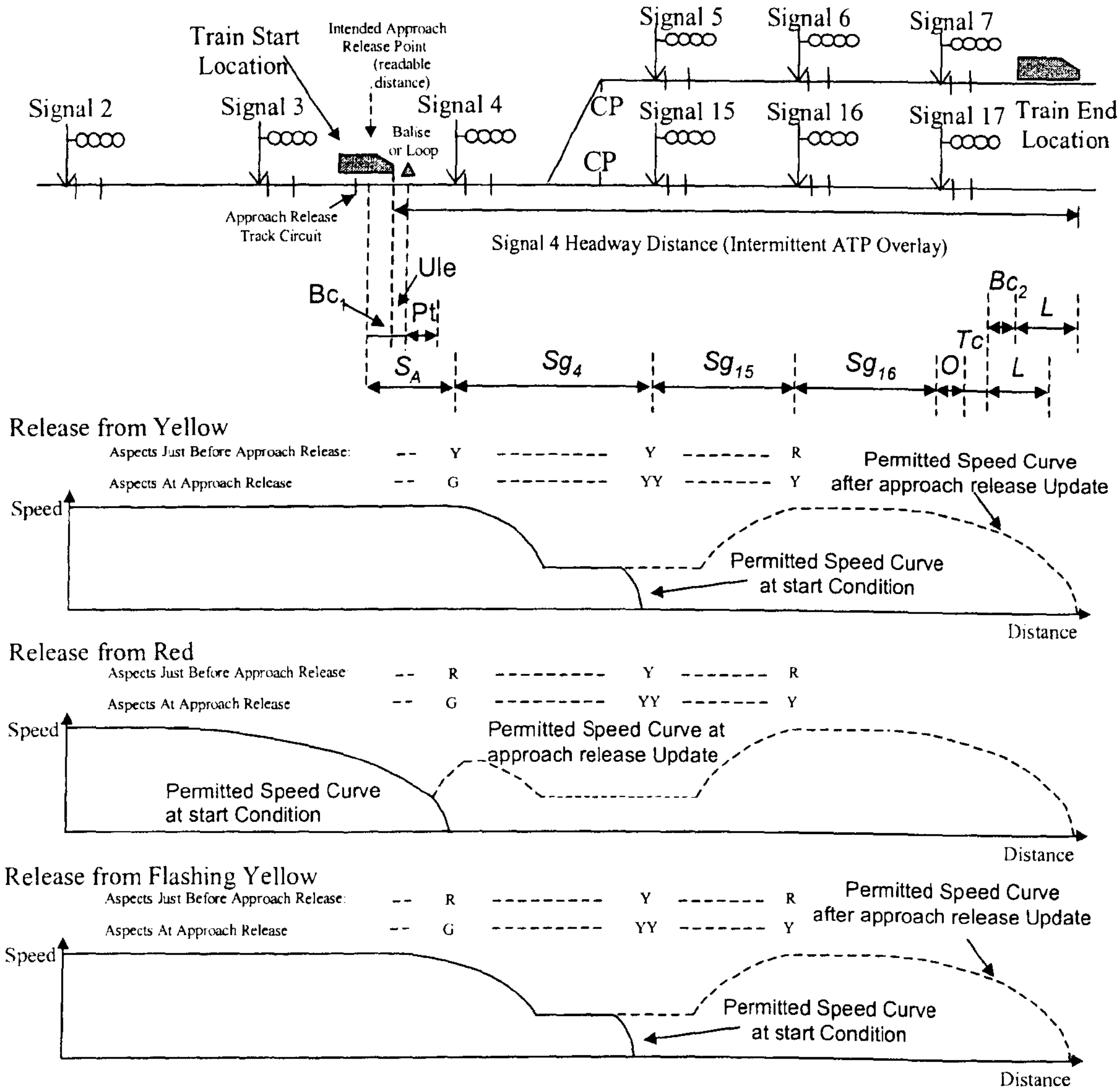


Figure F 2-6: Intermittent ATP Overlay Headway for Consecutive Diverging Trains with Approach Release

F2.1.5 CONSECUTIVE DIVERGING TRAINS - CONTINUOUS ATP OVERLAY

F2.1.5.1 NO APPROACH RELEASE OR SPLITTING DISTANT SIGNALS

As in the cases of conventional signalling and intermittent ATP overlay, the headway of consecutive trains taking the diverging route at a junction with no approach release or splitting distant signals is the same as that for a plain line section with an equivalent speed profile. The headway time for this scenario can be calculated using equation 9-2.

F2.1.5.2 APPROACH RELEASE FROM YELLOW, RED OR FLASHING DOUBLE YELLOW

As in the case of an intermittent update of supervision criteria, the trainborne supervision equipment of a continuous ATP overlay system will be unaware of the change in signal aspect that occurs at the planned clearance point (and will therefore continue to enforce the braking curve appropriate for the controlled aspect) until it receives and processes updated supervision criteria. Therefore, the

ATP 'sighting' point will be a distance equivalent to the message update delay (U_t) after the underlying signalling system's planned clearance point. The headway distance will then extend a distance equivalent to the message update delay after the end of the conventional signalling headway distance. This scenario is represented in Figure F 2-7.

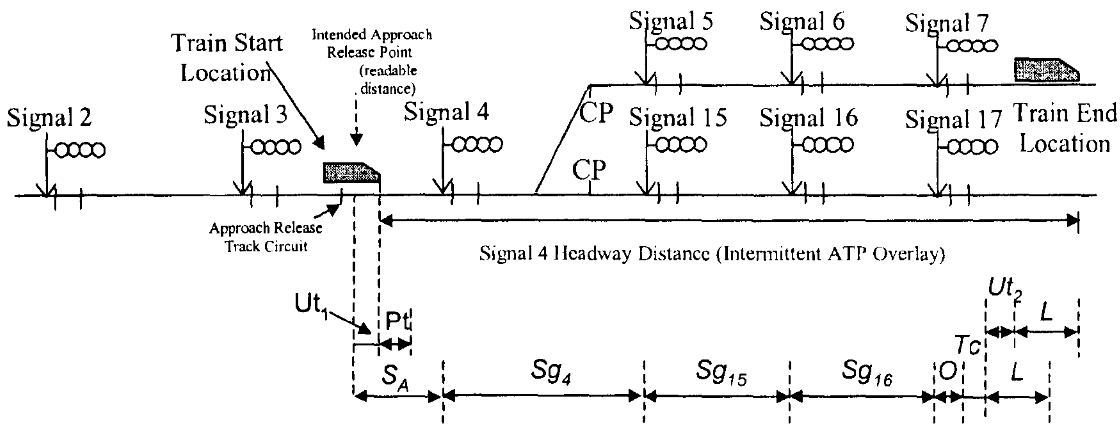


Figure F 2-7: Continuous ATP Overlay Headway for Consecutive Diverging Trains with Approach Release

It should be noted from Figure F 2-7, that due to the variation in operating speed through the line, the distances travelled by the lead and following train during the supervision update time (shown as Ut_1 and Ut_2) may differ. The headway time under Continuous ATP overlay operation is given by:

$$Ht_{CATP-O} = \sum_{x_{start}}^{\left(x_{start}+(S_A-Ut_1)+O+Tc+Ut_2+L+\sum_{p=4}^{p=n+3} Sg_p\right)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation F 2-5}$$

Where x_{start} is the location at which the trainborne equipment should receive transmission of updated supervision criteria following approach release of the junction signal (Ut_1 beyond the intended clearance point of the underlying signalling junction signal).

The forward speed profile used to calculate $V_{act}(x)$ must again represent the speed profile that the driver would have to follow to pre-empt intervention by the ATP system and comply with both the lineside signal aspects and professional driving practice. This would vary with the scenarios, as indicated by the speed / distance curves shown in Figure F 2-6 and the discussion on professional driving practice in section F2.1.1, since the supervision curve desired would not change between intermittent and continuous update. Once again, the train will be forced to follow the ATP system's restrictive supervision curve (based on the controlled aspect) until the updated supervision criteria have been received and processed on-board. This means that the train's speed will be restricted by the ATP system for the duration of the combined message update delay and trainborne processing delay (Ut+Pt) beyond the underlying signalling system's planned clearance point.

Whilst the distance covered by the trains may differ between Ut_1 and Ut_2, the times determining those distances are the same. If the three distances within Equation F2-5 (namely Ut_1, Ut_2 and Tc) are removed from the summation and instead added as their natural time elements, the two Ut values cancel and the equation becomes identical to Equation 8-14.

F2.1.6 CONSECUTIVE DIVERGING TRAINS - CONTINUOUS IN-CAB ATP

Since lineside signals are not required with continuous in-cab displays, the headway distance is always as shown in Figure F 2-8 and there is no need to consider the effects of approach control. With continuous in-cab ATP, the headway of consecutive trains taking the diverging route at a junction is therefore the same as that for a plain line section with an equivalent speed profile. The headway time for this scenario can be calculated using equation 9-3.

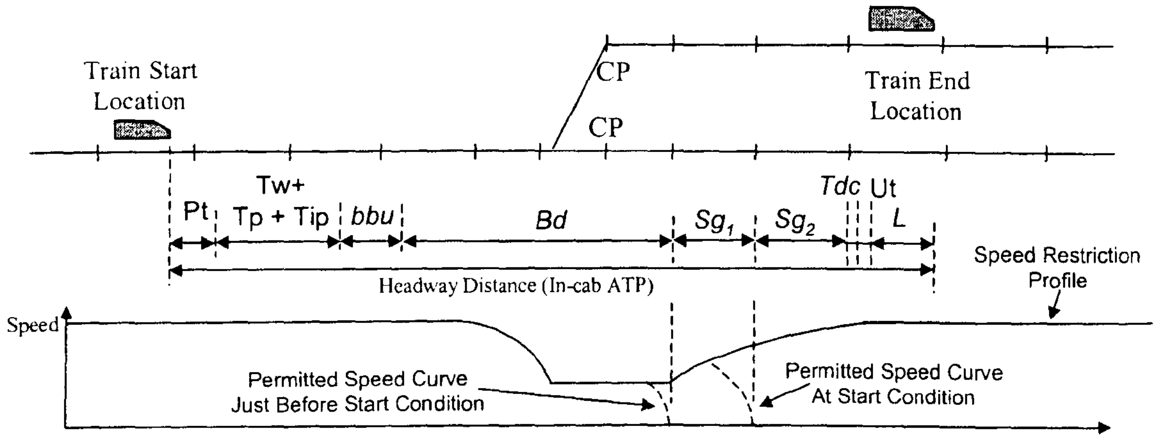


Figure F 2-8: In-Cab ATP Headway for Consecutive Diverging Trains

The longest headway time for the scenario will be the one including the lowest speed elements of the speed profile. The track section end that this leads up to (Sg_1 in Figure F 2-8) will depend on the particular speed profile and the extent of the track sections around the junction being considered.

F2.2 CONVERGING JUNCTION HEADWAY

In sections 7.3.6 and 8.4, the equation for converging junction headway was shown to be:

$$H_t = \textit{Train Following Headway} + \textit{Converging Train Run Time} - \textit{Straight Through Train Run Time} \qquad \text{Equation 7-36}$$

As can be seen by the example of a straight through train followed by a converging train shown in Figure F 2-9, this applies equally where an ATP system has been implemented. Whilst the primary example given is for an intermittent ATP overlay, the plain line headway distances shown in the speed distance profile of Figure F 2-9 highlight the minimal differences that would exist in the case of continuous overlay. In both cases, the headway time through a converging junction varies from that of a conventional signalling system only by the plain line headway. Hence, the headway time equations for the these two types of ATP system would be given by:

$$H_{IATP\ O} = \left(\sum_{x_{start}}^{(x_{start}+Ule+Pt+Tw+Tp+Tip+bbu+Bd+Sg+O+Tc+Bc+L)} \frac{\Delta x}{V_{act(lead)}(x)} \right) + \left(\sum_{SP_{conv}}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv)}(x)} \right) - \left(\sum_{SP_{str}}^{(FS_{conv})} \frac{\Delta x}{V_{act(str)}(x)} \right)$$

Equation F 2-6

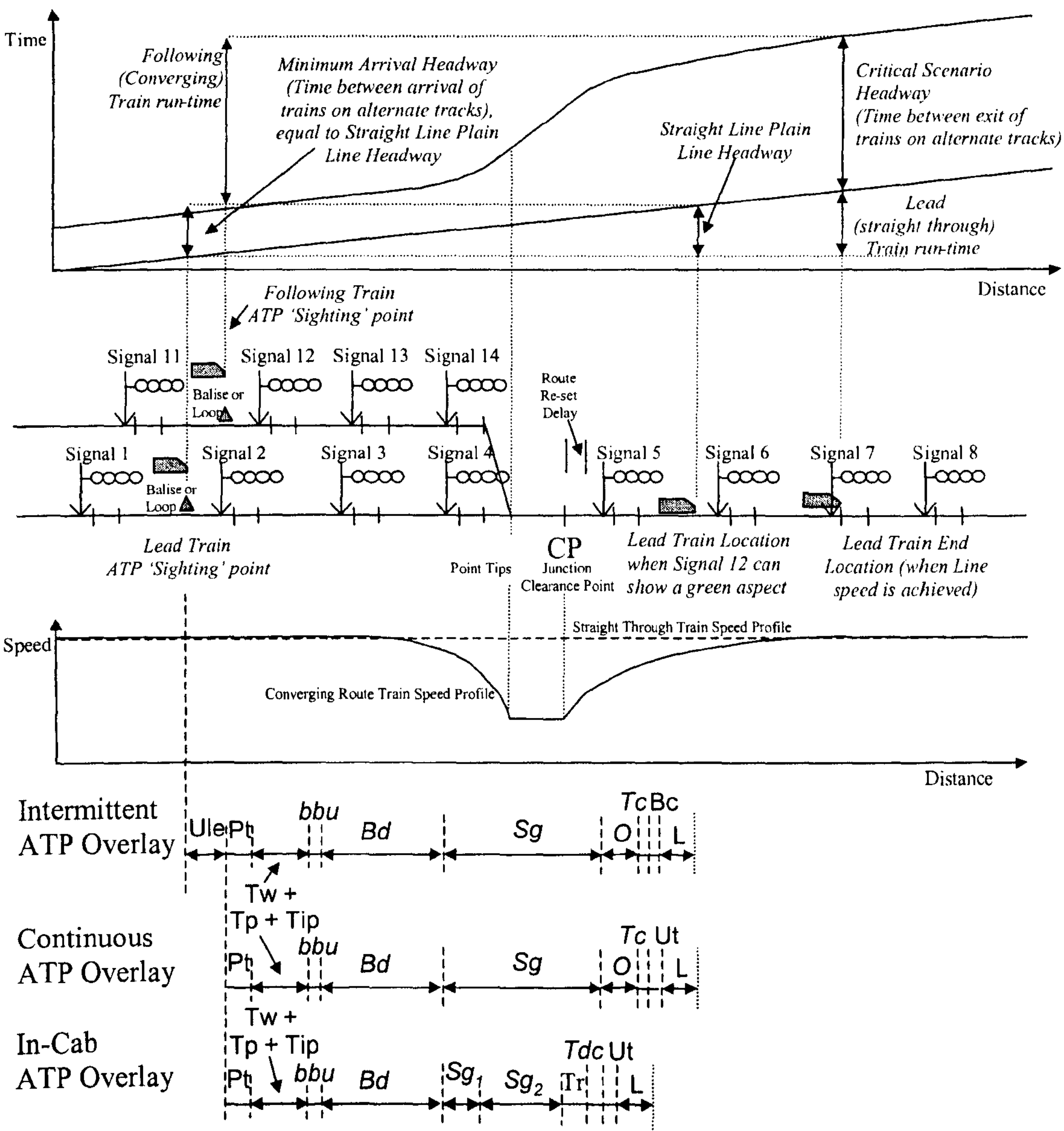
$$H_{CATP\ O} = \left(\sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+Sg+O+Tc+Ut+L)} \frac{\Delta x}{V_{act(lead)}(x)} \right) + \left(\sum_{SP_{conv}}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv.)}(x)} \right) - \left(\sum_{SP_{str}}^{(FS_{conv})} \frac{\Delta x}{V_{act(str.)}(x)} \right)$$

Equation F 2-7

Where: The converging train run time is taken to be the time from the effective ATP 'sighting' point of the first junction protection signal on the converging line (SP_{conv}) to the point at which a converging train attained the maximum permitted speed, having traversed the junction (FS_{conv});

The straight through train run time is taken to start at the effective sighting point of the first junction protection signal (SP_{str} on the straight through line) and to end at FS_{conv}, although it could, in practice, be at any point after FS_{conv};

The 'follow' train is whichever of the converging and straight through trains is last to reach



its 'SP'.

Figure F 2-9: ATP Headway for Converging Junction (Primary Example For Intermittent Overlay)

The case of in-cab ATP is a little different. Since the ATP is no longer constrained by signal locations, the block sections used can be much smaller and the point locking track section can be used as a block section in its own right. This significantly reduces the headway distance that would be required between consecutive trains travelling in the same direction and makes the approach headway for converging trains dependent on the junction route reset, as shown in Figure F 2-9. The headway time equation for a straight through followed by converging train at a converging junction signalled with in-cab fixed block is given by:

$$H_{CATP-In} = \left(\sum_{x_{start}}^{(x_{start}+Pt-Tw+Tp+Tip+bbu+Bd+Sg_1+Sg_2+Tdc+Tr+Ut+L)} \frac{\Delta x}{V_{act(lead)}(x)} \right) + \left(\sum_{SP_{conv}}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv)}(x)} \right) - \left(\sum_{SP_{str}}^{(FS_{conv})} \frac{\Delta x}{V_{act(str)}(x)} \right)$$

Equation F 2-8

The headway time for a converging train followed by a straight through train would not be dependent on the junction reset time and would, therefore, be given by:

$$H_{CATP-In} = \left(\sum_{x_{start}}^{(x_{start}+Pt-Tw+Tp+Tip+bbu+Bd+Sg_1-Sg_2+Tdc+Ut+L)} \frac{\Delta x}{V_{act(lead)}(x)} \right) + \left(\sum_{SP_{conv}}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv.)}(x)} \right) - \left(\sum_{SP_{str}}^{(FS_{conv})} \frac{\Delta x}{V_{act(str.)}(x)} \right)$$

Equation F 2-9

F2.3 MULTIPLE PLATFORM STATION HEADWAY

The headway through a station with a single in line platform per direction of travel can easily be calculated with the equations already derived for plain line headway, simply by accounting for the variation in $V_{act}(x)$ (including allowance for a dwell period). However, the case of multiple platforms warrants further consideration with the introduction of ATP.

In section 8.5.2, it was noted that the use of multiple platforms requires the use of junctions between the main line and off-line platforms. It was further noted that the likely differentials between turnout and straight through speeds at these junctions would typically require approach release from red to be applied to the last home signal of an underlying signalling system protecting the diverging junction.

F2.3.1 ALL TRAINS STOPPING - ATP OVERLAY

In the discussion for conventional signalling in section 8.5.2.1, it was noted that the approach to a multiple platform station represents a more restrictive headway than departure from it (except in the case of 3-aspect signalling with low line speeds). The headway through a two platform per direction station signalled with 4-aspect signalling and an intermittent ATP overlay system would be as shown in Figure F 2-10, for which it has been assumed that:

- Consecutive trains enter alternate platforms;
- Train 1 takes the diverging (approach released from red) route to the off-line platform.

A detailed explanation of the identified headway distances can be found in section 8.5.2.1 and will not be repeated here. Whilst Figure F 2-10 specifically represents the case of an intermittent ATP overlay system, that of a continuous ATP overlay system would only differ in the update times (Ut instead of Bt) and absence of any Update location error (Ule).

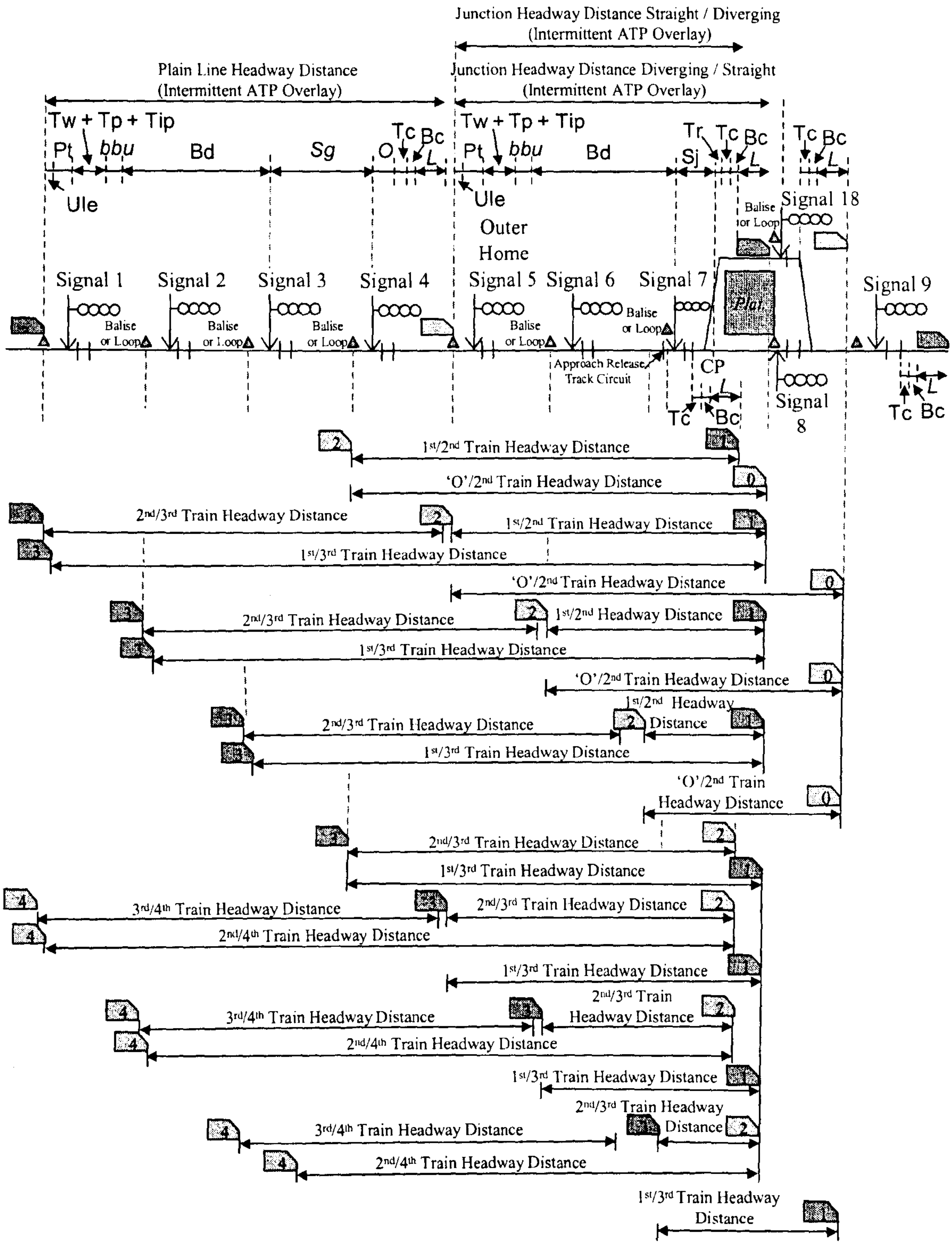


Figure F 2-10: 4-Aspect Headway with Intermittent ATP overlay, for Two Platforms per Direction of Travel (All Trains Stopping)

As with the case of conventional signalling without ATP (considered in section 8.5.2.1), the headway time required between the 1st and 2nd trains can be determined by considering the arrangement to be a diverging junction, with the longest headway through the junction being that of the outer home (or first junction protection) signal. For an intermittent ATP overlay system the headway times would, therefore, be given by Equation F 2-1 (Section F2.1.1.1) and for a continuous ATP overlay system by Equation F 2-2 (Section F2.1.2.1).

Since the inner home signal is approach released from red for diverging routes, the headway distance required between the 2nd and 3rd trains at the outer home will be smaller than that between the 1st and 2nd trains. However, the speed of travel over this approach-released route will be lower. Thus, the relative significance of the two headway times will be determined by the FSP of the respective routes. The headway time required between the 2nd and 3rd trains will be given by:

$$Ht_{\text{IATP-O, 2nd / 3rd}} = \sum_{x_{start}}^{(x_{start}+Ule+Pt+Tw+Tp+Tip+bbu+Bd+O+Tc+Bc+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation F 2-10}$$

$$H_{CATP-O,\text{2nd / 3rd}} = \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+O+Tc+Ut+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation F 2-11}$$

Whilst the first junction protection signal represents the longest headway through the actual junction, it should again be remembered that the plain line headway on approach must also be considered in order to determine whether the junction actually impedes overall headway.

For the critical headway condition between the 1st and 3rd trains, the headway time required between the 2nd and 3rd trains can be determined by considering the arrangement to be a plain line section. For an intermittent ATP overlay system the headway time is given by equation 9-1 (Section 9.2) and for a continuous ATP overlay system by equation 9-2 (Section 9.3). This would also be the case for the headway time required between the 3rd and 4th trains for the critical headway condition between the 2nd and 4th trains.

As shown in Figure F 2-10, on approach to the station area the headway distance required between alternate trains is determined by that required between consecutive trains. However, once the headway distance extends into the station area, where consecutive trains occupy different platforms, this is no longer the case. For the 1st and 3rd trains, the most significant headway that includes the station stop is that for the inner home signal, given by:

$$Ht_{IATP-O(Alernate\ trains,\ diverging,\ multiple\ platforms)} = Dwell + \sum_{x_{start}}^{(x_{start}+(S_A-Bc_1)+Sg_{ih}+O+Tc+Bc_2+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation F 2-12}$$

$$Ht_{CATP-O(Alernate\ trains,\ diverging,\ multiple\ platforms)} = Dwell + \sum_{x_{start}}^{(x_{start}+(S_A-Ut_1)+Sg_{ih}+O+Tc+Ut_2+L)} \frac{\Delta x}{V_{act}(x)} \qquad \text{Equation F 2-13}$$

Where 'ih' is the number of the inner home (7 in the example of Figure F 2-10);

Refer to section F2.1.4.2 for detailed explanation of the terms in these equations.

For the 2nd and 4th trains, the most significant headway that includes the station stop is that for the outer home signal, given by:

$$Ht_{\text{IATP-O}(Alternate\ trains,\ mainline,\ multiple\ platforms)} = Dwell + \sum_{x_{start}}^{(x_{start}+Ule+Pt+Tw+Tp+Tip+bbu+Bd+Sg_{ih}+O+Tc+Bc+L)} \frac{\Delta x}{V_{act}(x)}$$

Equation F 2-14

$$H_{CATP-O(Alternate\ trains,\ mainline,\ multiple\ platforms)} = Dwell + \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+Sg_{ih}+O+Tc+Ut+L)} \frac{\Delta x}{V_{act}(x)}$$

Equation F 2-15

Where 'ih' is the number of the inner home (7 in the example of Figure F 2-10).

These equations represent a specific instance of the general plain line headway equations derived in sections 9.2 and 9.3 (equation 9-1 and 9-2).

As already discussed in section 8.5.2.1, equivalent equations to those derived in this section could be developed for the effect of providing any number of platforms.

F2.3.2 ALL TRAINS STOPPING – IN-CAB ATP

With the introduction of in-cab ATP and removal of lineside signals, the headway through a multiple platform station can still be calculated by the same approach as applied to ATP overlay systems. However, the track circuits can be laid out in a more optimal manner and movement authorities can be updated on clearance of any track section. The headway through a two platform per direction station signalled with an in-cab ATP system is shown in Figure F 2-11, for which it has been assumed that consecutive trains enter alternate platforms.

It should be noted from Figure F 2-11 that there is no need for approach release with in-cab ATP, so the headway distances applicable to diverging and straight through trains are the same, although the associated headway times will differ in accordance with the routes' speed profiles.

The headway between consecutive trains will be in accordance with the junction headway considered in section F2.1.3 (equation F 2-3).

The critical headway condition between alternate trains (that is, the 1st and 3rd or 2nd and 4th trains) which travel into the same platform will be that for which the lead-train has just cleared the platform track section. This means that the headway time will be given by:

$$H_{CATP-In(TrainsToSamePlatform)} = Dwell + \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+Sg_p+Sg_s+Tdc+Ut+L)} \frac{\Delta x}{V_{act}(x)}$$ Equation F 2-16

Where: 'Sg_s' is the track section equivalent to the starter overlap (i.e. immediately beyond the platform). 'Sg_p'is the track section that contains the departure end of the platform.

This equation represents a specific instance of the general plain line headway equation derived in section 9.4 (equation 9-3). As already discussed in section 8.5.2.1, equivalent equations could be developed for the effect of providing any number of platforms.

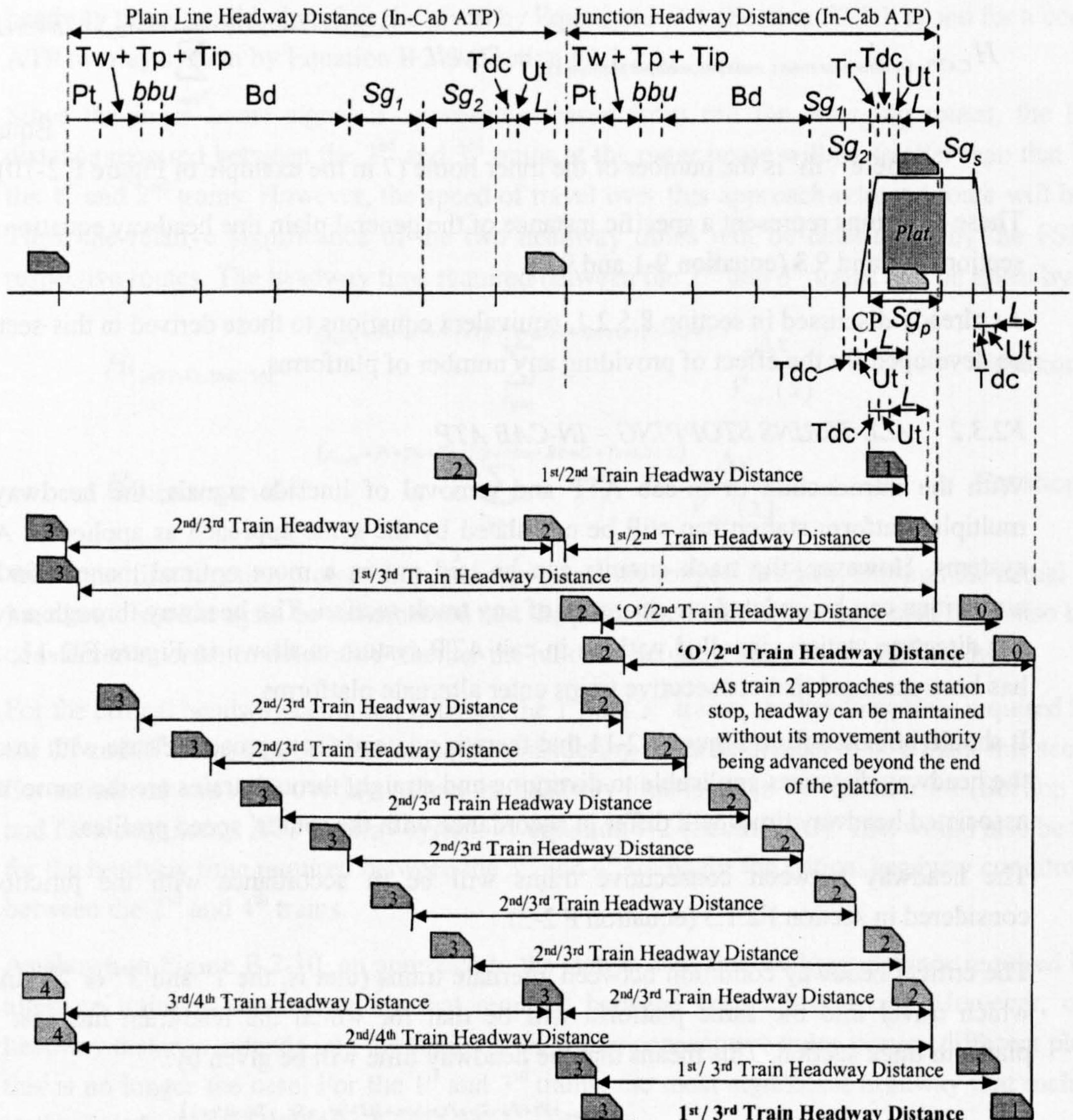


Figure F 2-11: Headway with In-Cab ATP, for Two Platforms per Direction of Travel (All Trains Stopping)

F2.3.3 STOPPING / NON-STOPPING COMBINATIONS WITH ATP

It was shown in section 8.5.2.2 that the headway on approach to an off-line platform can be treated as a diverging junction with approach control from red, whilst departure from it can be treated as a converging junction.

With the introduction of an ATP system, the scenario outlined in Figure 8-15 still applies and the critical headway time for both the stopping/non-stopping and non-stopping/stopping scenarios is still given by Equation 7-36. However, a few changes must be made when implementing this equation:

- The plain line train following headway component must change to reflect the impact of ATP delays and sighting points, in accordance with equations 9-1, 9-2 and 9-3;

- The train run-time components must commence at the effective ATP sighting points rather than the signal sighting points and;
- The station stop must be accounted for.

This results in a revised headway time equation, given by:

$$Ht_{n\,IATP-O} = Dwell + \left(\sum_{x_{start}}^{(x_{start}+Ulc+Pt+Tw+Tp+Tip+bbu+Bd+Sg+O+Tc+Bc+L)} \frac{\Delta x}{V_{act(lead)}(x)} \right) + \left(\sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv)}(x)} \right) - \sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(str)}(x)}$$

Equation F 2-17

$$Ht_{n\ CATP-O} = Dwell + \left(\sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+Sg+O+Tc+Ut+L)} \frac{\Delta x}{V_{act(lead)}(x)} \right) + \left(\sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv)}(x)} \right) - \sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(str)}(x)}$$

Equation F 2-18

$$Ht_{n\ CATP-In} = Dwell + \left(\sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+Sg_1+Sg_2+Tdc+Ut+L)} \frac{\Delta x}{V_{act(lead)}(x)} \right) + \left(\sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv)}(x)} \right) - \sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(str)}(x)}$$

Equation F 2-19

Where SP is the Sighting point of the ATP overlay system and FS_{conv} is the point at which a converging train would attain the maximum permitted speed, having travelled through the station area.

As with the case of all trains stopping, the headway that can be achieved between consecutive trains is not necessarily the most significant limitation. Before a stopping train can be routed into the off-line platform, any preceding train with the same routing must have cleared the platform. The equations required to consider this have already been derived in sections F2.3.1 and F2.3.2.

F2.4 TERMINAL STATIONS

The ways in which terminal stations differ from through station arrangements have already been considered in section 8.5.3. With the introduction of ATP, the headway to be expected on entry to a terminal station can still be calculated by treating it as travelling over a diverging junction with no approach release. The most restrictive headway will also still be encountered at either the first junction protection signal (the outer home signal) or the preceding signal, depending on the combination of forward speed profiles, junction location and route-reset delays.

The headway for the outer home signal can be determined by use of the equation already derived for the first junction protection signal of a diverging junction with no approach release. See equation F 2-1 (intermittent ATP overlay), equation F 2-2 (continuous ATP overlay) and equation F2-3 (in-cab ATP). In the overlay cases, also see equation 8-12 if the ATP sighting point is only required to be well within the sighting point of the underlying signalling, such that the ATP does not act as a constraint on achievable headway.

The headway of the preceding signal can be found by applying the terminal speed profile to the equations for train following headway. See equations 9-1 (intermittent ATP overlay), 9-2 (continuous ATP overlay) and 9-3 (in-cab ATP). As discussed in section 8.5.3, the same equations can be used to determine the headway between consecutive arrivals to different platforms for:

- Three or more terminal platforms fed by a single approach track (including cases where the approach track splits into two or more feeder tracks in advance of the platforms);

- Terminal stations provided with extended overrun tracks beyond the platform ends.

The minimum station exit headway with an ATP system can be calculated in the same way as that for conventional signalling (outlined in section 8.5.3.2), such that:

$$Ht_{n\,IATP-O} = \left(\sum_{Bl}^{(Bl+Bj+Sg_1+Tc+O+L)} \frac{\Delta x}{V_{act(first\_train)}(x)}\right) + \left(\sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(second\_train)}(x)}\right) - \sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(first\_train)}(x)}$$

Equation F 2-20

$$Ht_{n\,CATP-O} = \left(\sum_{Bl}^{(Bl+Bj+Sg_1+O+Tc+Ut+L+Pt)} \frac{\Delta x}{V_{act(first\_train)}(x)}\right) + \left(\sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(second\_train)}(x)}\right) - \sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(first\_train)}(x)}$$

Equation F 2-21

$$Ht_{n\,In-cabATP} = \left(\sum_{Bl}^{(Bl+Bj+S_1+S_2+Tr+Tdc+Ut+L+Pt)} \frac{\Delta x}{V_{act(first\_train)}(x)}\right) + \left(\sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(second\_train)}(x)}\right) - \sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(first\_train)}(x)}$$

Equation F 2-22

A few slight variations in the calculation that have been introduced as a result of the ATP system should be noted:

- In the case of an Intermittent ATP system, the following train must be able to begin moving when the signal clears, but must not pass the signal balise or loop until it has been updated with the new message. This acts as a constraint on the balise/loop location;
- In the case of a continuous or in-cab ATP systems, ATP updates can be issued before the train starts moving. In order for this to happen, the message update / transmission delay and train processing time must have elapsed;
- In the case of in-cab ATP, the train can be given a movement authority to depart as soon as the preceding train has cleared the first two track sections after the berth track section and any junction points in between the two trains have been set to the positions required by the following train. In the example given in Figure 8-16, this would make the point re-setting the last factor required for the following train to depart. The equation has been developed to reflect this, with the distance 'S_2' representing the length of the point locking track section and 'S_1' the length of the track section between the berth and point locking tracks.

The station exit headway must be at least equal to plain line headway. If the equation quoted above for in-cab ATP permits trains to depart with separations that are smaller than this, the plain line headway outside of the station will form the overriding constraint on departure intervals.

F3 MOVING BLOCK TRAIN CONTROL SYSTEM CAPACITY

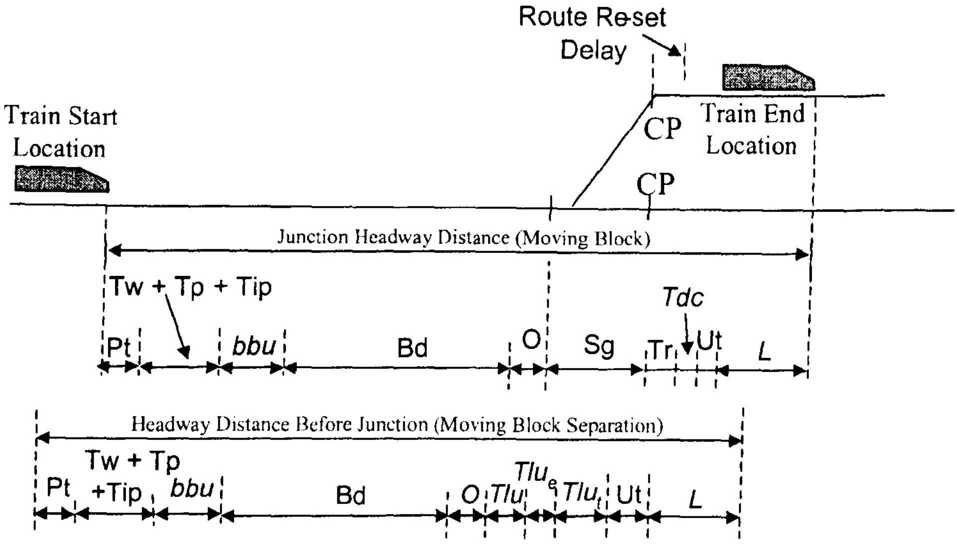
F3.1 DIVERGING JUNCTION HEADWAY

The introduction of junctions applies a fixed, location based, constraint to moving block headways. The headways for a diverging junction under conventional signalling or fixed block ATP system control have already been considered in chapter 8 (section 8.3) and section F2.1. In this section, similar consideration will be given to the impact of moving block.

Throughout the discussions in this section it will be assumed that the safety critical nature of point location and locking will necessitate the retention of point locking track circuits or axle counters.

F3.1.1 DIVERGING FOLLOWED BY STRAIGHT THROUGH TRAINS

As represented in Figure F 3-1, the junction point locking track section must be cleared by the leading (diverging) train before the route can be reset behind it for a following movement in the straight through direction. The overlap element of the train behind's movement authority cannot be



advanced beyond the start of this track section until the reset has occurred.

<u>Figure F 3-1: Moving Block Headway for Diverging Followed by Straight Through</u>

The junction headway time is given by:

$$Hm = \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+O+Sg+Tr+Tdc+Ut+L)} \frac{\Delta x}{V_{act}} \qquad \text{Equation F 3-1}$$

If it is considered that the distance 'O + Sg' is equivalent to the fixed block distance 'Sj' (the distance between the last junction protection signal and junction clearance point, it can be seen that equation F 3-1 is in-fact the same as equation F2-3 (the continuous in-cab ATP junction headway for a diverging / straight through train combination). This means that the headway for this scenario under moving block operation would be the same as that under fixed block in-cab ATP operation.

F3.1.2 CONSECUTIVE DIVERGING TRAINS

With consecutively diverging trains, the lie of the points does not need to change and the point locking track section will thus not form a constraint on the operation of the system. As a result, the headway of consecutively diverging trains over a junction under moving block control is equivalent to that of a plain line section with the same speed profile (given by Equation 9-4).

F3.2 CONVERGING JUNCTION HEADWAY

In sections 7.3.6 and 8.4, the equation for fixed block converging junction headway was shown to be:

$$H_t = \textit{Train Following Headway} + \textit{Converging Train Run Time} - \textit{Straight Through Train Run Time} \qquad \text{Equation 7-36}$$

As can be seen by the example of a converging junction shown in Figure F 3-2, this equation applies equally for a converging train followed by a straight through train where a moving block system has been implemented.

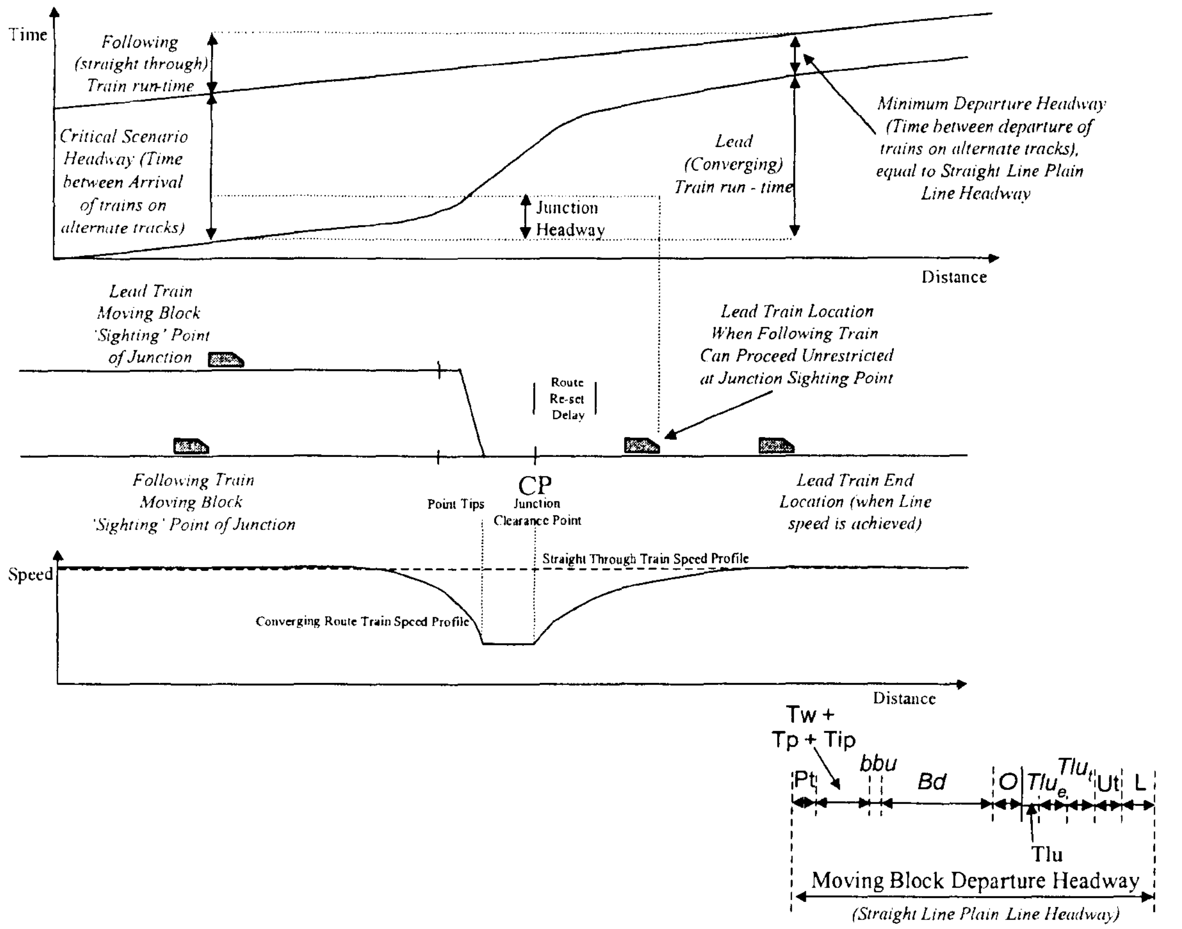


Figure F 3-2: Moving Block Headway for Converging Junction: Converging Followed by Straight Through

The moving block headway time for a converging train followed by a straight through train is given by:

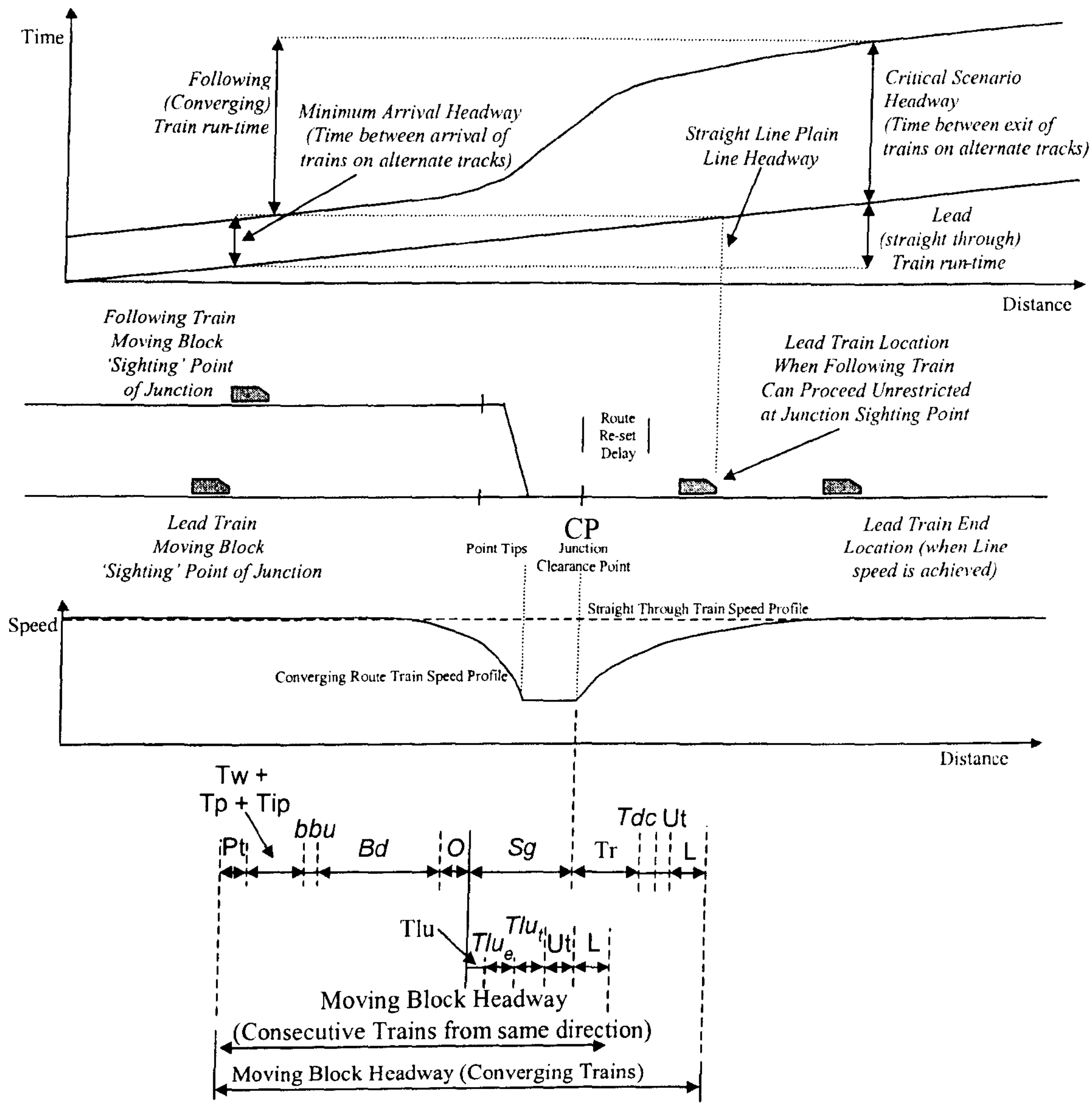
$$H_M = \left(\sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+O+Tlu+Tlu_e+Tlu_t+Ut+L)} \frac{\Delta x}{V_{act(lead)}(x)} \right) + \left(\sum_{SP_{conv}}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv.)}(x)} \right) - \left(\sum_{SP_{str}}^{(FS_{conv})} \frac{\Delta x}{V_{act(str.)}(x)} \right)$$

Equation F 3-2

Where: The converging train run time is taken from the effective moving block 'sighting' point of the junction on the converging line (SP_{conv}) to the point at which a converging train attained the maximum permitted speed having traversed the junction (FS_{conv});

The straight through train run time is taken to start at the effective moving block sighting point of the junction on the straight through line (SP_{str}) and to end at FS_{conv}, although in practice it could be at any point after FS_{conv}.

As with in-cab ATP, a moving block system is not constrained by signal locations. It is also not constrained by track sections (except for point locking track sections). This significantly reduces the headway distance that would be required between consecutive trains travelling in the same direction and makes the headway for a straight through train followed by a converging train dependent on the



junction route reset, as shown in Figure F 2-9.

Figure F 3-3: Moving Block Headway for Converging Junction: Straight Through Followed by Converging

Hence, the moving block headway time equations for a straight through followed by a converging train would be given by:

$$H_m = \left(\sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+O+Sg+Tdc+Tr+Ut+L)} \frac{\Delta x}{V_{act(lead)}} \right) + \left(\sum_{SP_{conv}}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv\_train)}} \right) - \left(\sum_{SP_{str}}^{(FS_{conv})} \frac{\Delta x}{V_{act(str\_train)}} \right)$$

Equation F 3-3

F3.3 MULTIPLE PLATFORM STATION HEADWAY

F3.3.1 ALL TRAINS STOPPING

The headway through a multiple platform station can still be calculated by the same approach as that applied to fixed block ATP systems. The headway through a two platform per direction station signalled with an in-cab ATP system would be as shown in Figure F 2-11, where it has been assumed that consecutive trains enter alternate platforms.

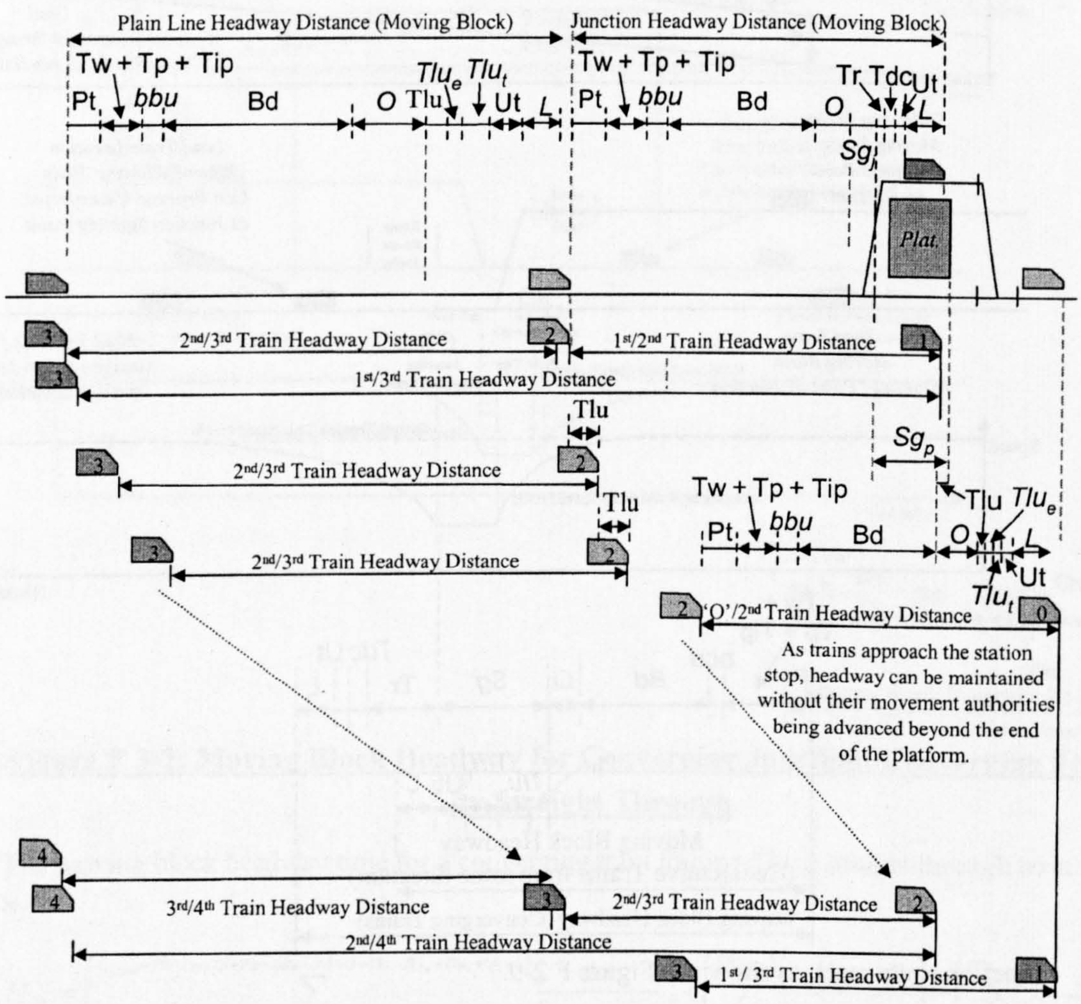


Figure F 3-4: Moving Block Headway For Two Platforms per Direction of Travel (All Trains Stopping)

The headway between consecutive trains travelling through the junction will be in accordance with the junction headway considered in section F3.1.1 (equation F3-1), with 'Sg' being replaced by the junction point locking track section, 'Sj'.

The critical headway condition between alternate trains (that is, the 1st and 3rd or 2nd and 4th trains) which travel into the same platform, will be that for which the lead-train has just cleared the platform berth (that is, the area of the platform in which trains stop at the station). This means that the headway time will be given by the general plain line headway equation derived in section 9.5 (equation 9-4), with the braking distance calculated to end a distance equivalent to that travelled in the location update time (Tlu) before the end of the platform section of track (Sg_p). This represents the situation when the updated movement authority extends to the end of Sg_p, allowing the train to proceed unrestricted into the station.

F3.3.2 STOPPING / NON-STOPPING COMBINATIONS

With the introduction of a moving block control system, the headway of a stopping/non-stopping service combination can still be calculated by the approach laid out in sections 8.5.2.2 and F2.3.3. Allowing for the differences that will be experienced in plain line train following headway, the headway time equation becomes:

$$Ht_m = Dwell + \left(\sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu+Bd+O+Tlu+Tlu_e+Tlu_t+Ut+L)} \frac{\Delta x}{V_{act(flead)}} \right) + \left(\sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(conv\_train)}} \right) - \sum_{SP}^{(FS_{conv})} \frac{\Delta x}{V_{act(str\_train)}}$$

Equation F 3-4

Where SP is the 'Sighting Point' of the moving block system on approach to the diverging junction before the station and FS_{conv} is the point at which a converging train would attain the maximum permitted speed having travelled through the station area.

As with the case of all trains stopping, the headway that can be achieved between consecutive trains is not necessarily the most significant limitation. Before a stopping train can be routed into the off-line platform, any preceding train with the same routing must have cleared the platform. The equations required to consider this have already been derived in section F3.3.1.

F3.4 TERMINAL STATIONS

With the transition from fixed to moving block, the headway to be expected on entry to a terminal station can still be calculated by treating it as travelling over a diverging junction.

The headway through the junction arrangement at the terminal throat can be determined by use of equation 3-1, derived for a diverging followed by straight through train in section F3.1.1. Although consecutive trains may diverge to different platforms, they will not go to the same one. Therefore this will apply to one set of points on approach to the station for each possible destination pairing). The headway on approach to the junction diverting trains to different platforms can be found by applying the terminal speed profile to the equations for plain line train following headway (see equation 9-4 in section 9.5).

The minimum station exit headway with a moving block ATP system can be calculated in the same way as that for in-cab ATP (outlined in section F2.4), however, in the absence of track sections the equation must be re-written:

$$Ht_m = \left(\sum_{Bl}^{(Bl+Bjc+Tr+Tdc+Ut+L+Pt)} \frac{\Delta x}{V_{act(first\_train)}} \right) + \left(\sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(second\_train)}} \right) - \left(\sum_{Bl}^{(FS_{conv})} \frac{\Delta x}{V_{act(first\_train)}} \right)$$

Equation F 3-5

Where: Bjc is the distance between the second train's berth location in the platform and the departure junction clearance point, i.e. the equivalent distance to $Bj+S_1+S_2$ in the fixed block in-cab signalling equation F 2-22.

The station exit headway must be at least equal to plain line headway. If the equation quoted above for in-cab ATP permits trains to depart with separations that are smaller than this, the plain line headway outside of the station will form the overriding constraint on departure intervals.

F3.5 SIMPLIFIED / TARGETED MOVING BLOCK

In his deliberations up to this point, the author has assumed the application of a comprehensive pure moving block system throughout a section of line. However, since the early days of moving block two simplified forms have also been suggested: moving space block and moving time block. To the author's knowledge, these arrangements were first proposed in three PhD theses, the first by a Loughborough University student in the 1970s (Pearson 1973, p5) and the other two by Birmingham University students during the following two decades (Gill 1986, pp4/24-6; Hiroto 1997, pp2/5 to 2/6 and Appendix F). The latter two of these sources clearly derive their inspiration from the former.

Under moving space block, trains are supervised against the maximum speed profile for the line and the braking distance from that speed, rather than the actual speed, is maintained between trains. The permitted speed is reduced to zero and the brakes must be applied when a train reaches a location at which braking from this maximum speed is required in order to stop the train a safety margin before the last reported location of the preceding train, or any fixed point being protected (Gill 1986, p4/24; Hiroto 1997, pp2-5 to 2-6; Pearson 1973, p5).

Hiroto states that moving space block is the simplest scheme for implementation of a moving block, since "the only information which the following train requires under moving space block is the position of the leading train in relation to its own position or any applicable limit of movement authority" (Hiroto 1997, p2-5). The implication in this statement is that moving space block minimises the amount of dynamic information that the system must know, since speeds are not important. However, this is not really true, since position and speed information are generally derived from the same source data (typically tachometers, accelerometers, etc.), and the speed information is still required on-board the train for the purposes of driver displays. It is purely in the calculation of braking curves that implementation of this arrangement might allow some reduction in processing requirements (with less variables to consider). With the processing power available in the 1970s this may have been a significant consideration, but the advantage appears fairly minimal in the light of modern technology. The arrangement also significantly reduces the potential for optimisation of intermixed traffic operation and improved recovery following perturbation when compared with full moving block. Indeed, Hiroto's analysis found that the number of trains perturbed by a single delay to one train under moving space block was similar to that for fixed block operation (Hiroto 1997, p4-15). In the light of these findings and considerations, the case for considering this arrangement further appears very weak and the author will not pursue it further.

Under moving time block, trains are given a target speed equivalent to the line speed as far as the location at which braking from line speed would be required in order to stop the train a safety margin before the last reported location of the preceding train, or any fixed point being protected. The maximum permitted speed is then reduced to keep the time separation between trains constant throughout the line (Hiroto 1997, p2-6; Pearson 1973, p5). In practice, the inclusion of a safety margin would mean that whilst the braking distance separation could be maintained at a constant equivalent time, the time separation would actually increase with reducing speed. However, accepting that this gives a headway time that is approximately constant, the benefit of such an implementation still needs to be explained. Unfortunately, neither Gill, Hir0to nor Pearson suggest what this might be. On further analysis, it is clear that the arrangement offers improved perturbation recovery when compared with moving space block, by allowing trains to close up at reduced speeds, as was found by Hiroto in his study of perturbation recovery (Hiroto 1997, pp7-2 to 7-3). It would similarly offer some improvement for optimisation of intermixed traffic operation. However, full moving block allows for both of these to a greater degree. In consequence, the case for further analysis of this arrangement again appears very weak and the author will not pursue it further.

In addition to these alternative moving block arrangements, Pearson also proposed that a 'hybrid system' could be implemented, where "the criterion for separating the trains might change according to conditions" (Pearson 1973, p7). Pearson's proposal was actually for switching between the three moving block forms. Considering the complexity that would be involved in such an arrangement, as well as the dubious benefits of the two 'simplified' forms, Pearson concluded, "It is difficult to envision what extra advantage would be derived from such a system" (Pearson 1973, p7). Whilst the author would agree with him for the arrangements that he proposed, an alternative hybrid has been more recently proposed, which does appear to offer some potential benefit by also allowing for alternative methods of implementing moving block.

F3.5.1 OVERLAYED MOVING BLOCK

In 1998, the authors of an IRSE paper on metro signalling and Operations noted that moving block requires complex engineering and high criticality processes. In order to prevent catastrophic failure of the overall signalling system due to the loss of some element of the complex system (such as the radio communication between trains), they suggested that careful consideration should be given to graceful degradation. In particular, the paper's authors proposed that moving block should be applied as an overlay on a more conventional fixed block based ATP system. In the event of a moving block component failing, traffic could then continue to operate under the control of the fixed block signalling. The authors then developed this proposal by suggesting that the moving block system could be applied only where increased performance was required, with the remainder of the line operating on fixed block ATP. They further noted that the moving block functionality could be provided only in station areas, which have the most critical headway requirements (White et al. 1998, pp4, 6-7).

This suggestion was noted by Mr Lewis of Alstom Signalling, who proposed a method for simplified implementation of moving block in critical areas. Whilst most current moving block systems are based on each train determining its own location and transmitting this to trackside equipment, Lewis proposed a system using propagation delays to establish the relative distance between consecutive trains (Lewis 2001, p1). In principle, this proposal is similar to the approach adopted by the GE AATC moving block system installed on BART[1]. However, Lewis envisaged a more localised system in which train separation, rather than the distance to fixed trackside locations, would be determined on-board the following train. The proposal requires a transponder with an accurate time stamp at the rear of each train, and a receiver on the front. Each train could then

[1] (In the GE AATC system, trackside radio units transmit time stamped signals that are received by trainborne units, compared with the time of receipt and translated into distance measurements – see Appendix C for more details)

receive signals from other trains (either by free air transmission or via a wave guide with known propagation characteristics), calculate the delay that occurred during transmission and thus the separation. In the absence of a known separation, trains would operate on the basis of a conventional fixed block signalling system. Whenever the receiver on the second train detected the transponder of the train in front and determined the separation, a localised moving block control would be activated, allowing the ATP to permit entry into 'occupied' blocks without compromising safety, and thus minimising headway in critical areas (Lewis 2001, pp1 -3).

In 1986, a Birmingham University student included a single paragraph within his thesis suggestion that a moving block system could be implemented by mounting a radar device to the front of trains to monitor the position and speed of any train (or obstacle) ahead. The author of that thesis then rejected the idea on the basis that "for metro type applications, problems may arise with scattering in tunnels and loss of detection on sharp curves" (Gill 1986, p4/23). If used as the general train detection mechanism throughout a line, this would certainly be true. However, such an arrangement would be less likely to experience difficulties if used to provide the more localised functionality envisaged by Messrs Lewis and White.

Whilst determining movement authority directly on the basis of vehicle to vehicle communication would require the development of suitable interlocking, radio and ATP equipment, it does appear to offer a viable approach to localised moving block control that would not require complex trackside processing to monitor train locations and would automatically fall back to a fixed block arrangement in the event of equipment failure of an un-fitted train. The main question raised by the proposal is whether such a localised implementation would offer capacity benefits where they are most needed.

The equations already developed within this appendix provide the means to address this question by appropriate use of the system delay parameters. In the author's opinion, this proposal does represent a significant potential for capacity benefit at minimum cost and, therefore, warrants further analysis.

F4 POSSIBILITIES AND CAPACITY OF OPERATION WITH LESS THAN BRAKING DISTANCE SEPARATION

F4.1 INTRODUCTION

All of the fixed and moving block train separation strategies considered so far have allowed for at least the full service braking distance and a safety margin between the last known location of a train and the point at which a restrictive movement authority must be issued to any following train. This approach ensures safety by guaranteeing that a train operating within its last received movement authority will be able to stop under service braking even if the train ahead is stationary. However, for the majority of train movements the leading train is moving and will not, therefore, still be at its last known location by the time that a following train could arrive there.

In this section, the author will begin by considering the practicalities and capacity impact of utilising knowledge of a leading train's movement, speed and performance characteristics, to allow movement authorities to include reduced safety margins or even to be based on the relative braking distances of consecutive trains, rather than the absolute braking distance of the following train.

The author will then move on to consider an alternative approach to reducing train separation by utilising the differentials in service and emergency braking rates. On completion of this section, it will be possible to assess the impact of the approaches considered on achievable headway, by comparison of the equations derived with those previously derived in chapters 8 to 9 and earlier parts of this appendix.

F4.2 PREDICTIVE RESPONSE TO PRECEDING TRAINS (FIXED BLOCK)

The author has personal experience of designing fixed block colour light signalling arrangements with predictive response to preceding trains, for implementation on LUL. Whilst it is usual practice on LUL to provide station starter signals with a reduced overlap, sufficient for 35km/h, for obvious safety reasons this can only be done where there is no conflict between the full speed overlap and a train berthed at a signal ahead or a train movement on another route (LUL undated, p1). If these conditions are not met, a train approaching the station is held at the home signal until the preceding train has cleared the overlap, resulting in a significant impact on headway.

Where the headway impact caused by adopting an overlap length equivalent to full-calculated braking distance would make the starter signal's headway a significant bottleneck for the line, LUL practice allows the use of a speed checked 'creeper' signal on approach to the starter that has a full speed overlap ending before the conflict. The starter signal's overlap can then be based on acceleration from the checked speed between the creeper and starter signal, with a minimum overlap provision of 25km/h (LUL undated, pp1-2). This method for mitigating the risk of overrun requires provision of additional equipment, including a signal head in the platform area. In some locations this is not feasible, either due to space constraints or lack of spare contacts on existing relays that would have to be included in the creeper signal's selection circuits.

Where the use of a creeper is not feasible or desirable, two forms of predictive response are occasionally used instead. The first of these is known as 'Train Away', and is implemented for starter signals in automatically worked areas (such as the southbound starter signal at London Bridge). Under Train Away, the home signal is allowed to clear to a proceed aspect even if there is a train in the starter signal's full speed calculated overlap, so long as the reduced speed (35km/h) overlap is clear and the signal controlling the train ahead is also displaying a green aspect. These conditions are considered adequate to indicate that the preceding train is unlikely to stop (or remain stopped) within the full starter overlap and to predict that it will have cleared the overlap before the following train could pass the home signal and SPAD the starter signal at speed. A stick path in the home signal's selection circuit then ensures that it can remain at green once the leading train passes the signal that was controlling its movement (returning it to red).

Where signals can be controlled, and therefore returned to danger at any time, there is a risk that the conditions required for Train Away could be met, even though the signal after the starter could then be replaced to danger before the train approaching or berthed at it is able to proceed past. If this were to happen, Train Away would allow the home signal to continue displaying a green aspect with a stationary train held in the starter signal's full speed overlap – a potentially dangerous situation. In order to overcome this risk, a second technique known as 'Route Away' is implemented for starter signals in controlled areas (such as Stockwell and Euston). Route Away includes detection that a route from the signal after the starter has been set (rather than the signal being green), and that the signal's replacement track has been occupied (indicating that the leading train has passed the signal). When these conditions are met, with the reduced starter overlap track being clear, they provide a more robust indication that the train currently occupying a part of the starter signal's full speed overlap is moving off and the home signal is allowed to clear to a proceed aspect (as in Train Away).

In both cases, the LUL predictive response techniques are crude. In the case of Train Away, the signalling system does not know for certain that the leading train is moving through the route. It is just considered unlikely to stop on approach to a green signal. Similarly, whilst a train passing a green signal shows that it was moving at the time it passed, in the case of Route Away the signalling system still does not know how fast the train is moving or whether it will stop before clearing the full speed overlap. Therefore, whilst offering techniques to optimise headway through a pinch point on the line, they also introduce a significant amount of risk. In consequence, they are not techniques of choice when installing new signalling arrangements. They are instead used as partial mitigation techniques where existing arrangements are found to contain inadequate overlaps (generally due to less stringent requirements being in force at the original time of design – sometimes as long ago as the 1940s).

Another example of a fixed block system with predictive response to preceding trains can be found in proposals from the Railway Technical Research Institute in Japan. They suggested that a Digital ATP system could monitor the door open/close data and ATO profile of a train in a platform, in order to allow the following train to predict when the lead train would leave the platform and control its own approach in accordance with this prediction (Fukuda et al. 2002, pp5-6 and Figure 6d). Unfortunately, little detail of this feature was given in the paper describing the Digital ATP system.

As with the LUL Train Away and Route Away techniques, this proposal would only confirm that the conditions required for the lead train to depart had been met and it should ordinarily depart. It does not prove that the train has departed, or that it will not stop again before clearing the platform and a safety margin beyond it. The prediction being made is, therefore, subject to a high level of risk.

It is the opinion of the author that all three of these techniques for predictive behaviour under fixed block arrangements would be considered too risky for use on a new line or in a comprehensive re-signalling scheme. They offer a risk / capacity compromise for existing, outdated, systems, but no more than that. They have been discussed in this section because they show a precedent to considering the use of predictive response and, in the case of the LUL techniques, for its implementation. However, in light of the risks that they involve, the author does not intend to analyse their capacity impact any further.

F4.3 PREDICTIVE RESPONSE TO PRECEDING TRAINS (MOVING BLOCK)

With the introduction of a moving block arrangement, where each train determines and reports its own location to the central control system, it becomes possible to envisage an enhanced approach to predictive response. If trains reported their speed, and possibly even details of their braking performance and crashworthiness, the movement authority issued to a following train could be based upon the predicted stopping point of the leading train rather than its last known location. This approach could most simply be implemented as relative braking distance separation.

F4.3.1 RELATIVE BRAKING DISTANCE

The concept of operation with relative braking distance separation is based on the fact that a train cannot stop instantaneously in normal and most types of degraded operation. Therefore, if the speed and braking performance of the lead train are known, a movement authority can be given to the following train based on the earliest stopping point of the lead train, rather than its last reported location.

As a concept, relative braking has been around for quite some time. The earliest reference found by the author was a statement within a 1938 IRSE lecture, that "as a possible method of decreasing headway some form of cab signalling, which is dependent on the relative speeds of following trains, might be developed and be such that trains could always run with just sufficient braking distance between them" (Woodbridge 1938, p198). From the time of this suggestion, the concept has been subject to severe criticism. Mr Crook (then president of the IRSE) noted in the discussion that followed Mr Woodbridge's comments on relative braking that, if two trains were travelling at the same speed, "there need be no separating distance" (Woodbridge 1938, p204). Later authors have adapted the proposal slightly to overcome this particular objection (and also to allowance for unexpected variations in braking performance), such that "Spacing trains in relative braking distance means that the distance between two successive trains equals the difference of the braking distances of the trains plus an additional safety margin" (Pachl 2000, p3/1). However, a number of other objections still remain:

- If risk of derailment is to be avoided when points are moved between two trains, the second train must have full braking distance to the point tips until the points are locked again. This requires a separation greater than the relative braking distance of the two trains;
- Relative braking separation only precludes the compounding of collisions if no train decelerates at a higher rate than predicted. In the case of an accident causing the first train to stop at an exceptionally high rate, the second train has no chance to stop and is going to collide with the first train. In this way, multiple collisions of following trains could well be possible

  (Brauer 2001, p3; Pachl 2000, pp3/1-2; Pearson 1973, p7).

The first of these points is totally valid and implies that use of relative braking separation could not be made between trains taking different routes at a junction, unless some form of points without moving parts were to be invented (Brauer 2001, pp4-6). However, this restriction would not apply to trains approaching a single in line platform, where some benefit could potentially be gained by allowing trains to 'close up' as the leading train departs and the following train arrives.

The second objection is more serious and would apply to any occurrence of relative braking operation. In consequence of the potential safety implications of this concern, authors considering relative braking have variously concluded that it could never be implemented on a railway (Pachl 2000, pp3/1-2), would only be practicable for segregated freight traffic (Hiroto 1997, p2-8) or could only be applied for fully automated freight operations where all personnel have been removed from the vehicles (Pearson 1973, p7).

Despite this concern, early versions of the ERTMS System Requirements Specification included a chapter on the Radio Block Centre, which referred to the concept of relative braking distance and defined the choice between absolute and relative braking distance operation to be "a national decision" (ERTMS Users Group[2] 1998, p17). This clearly shows that the potential for operation with relative braking distance separation is being considered in Europe and that analysis of the capacity benefits that may be associated with it would, therefore, be of significant value.

The headway distances associated with relative braking operation can be seen in Figure F 4-1.

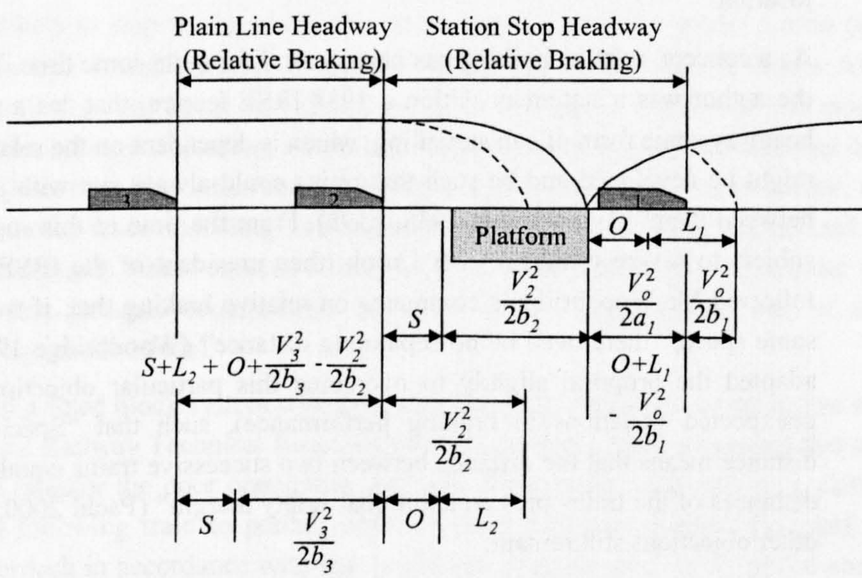


Figure F 4-1: Relative Braking Headways

The plain line headway time under relative braking operation is, therefore, given by:

$$H_{trb} = \left(\frac{V_f^2}{V_l \times 2b_f} - \frac{V_l}{2b_l} \right) + \frac{S + O + L_l}{V_l} \qquad \text{Equation F 4-1}$$

Where V_l and V_f are the maximum permitted speeds for the lead and following trains respectively.

The headway distance is converted into time by use of the lead train speed so that the result represents the time that will elapse between occupation of the start location by the lead train and the time that the following train can reach the same location whilst remaining headway distance behind the lead train. This is the headway time interval between the trains passing the start location.

Following a similar approach to determine the headway time between trains stopping in the platform gives the result:

$$H_{trb(Stopping)} = \frac{S + \frac{V_f^2}{2b_f} - \frac{V_l^2}{2b_l}}{V_l} + \frac{V_l}{b_l} + Dwell + \begin{cases} \sqrt{\dfrac{2\left(O + L_l - \frac{V_o^2}{2b_l}\right)}{a_l}} & \text{For } V_l \geq \sqrt{2a_l\left(O + L_l - \frac{V_o^2}{2b_l}\right)} \\ \dfrac{V_l}{2a_l} + \dfrac{O + L_l - \frac{V_o^2}{2b_l}}{V_l} & \text{For } V_l \leq \sqrt{2a_l\left(O + L_l - \frac{V_o^2}{2b_l}\right)} \end{cases}$$

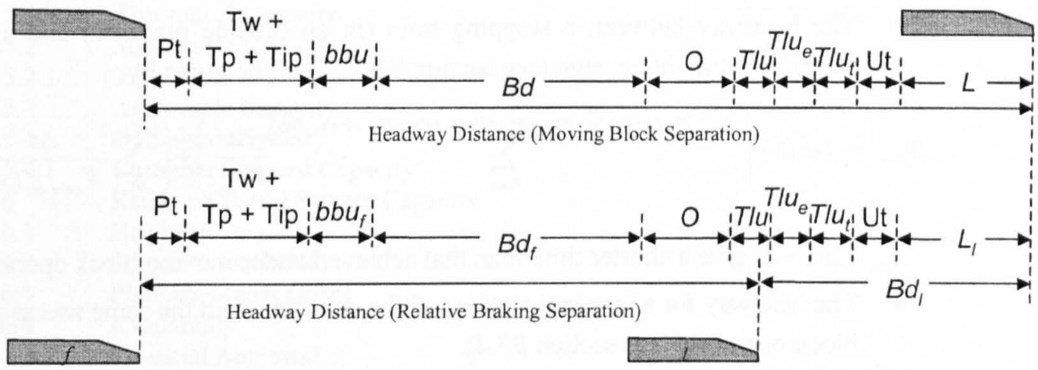
Equation F 4-2

Where V_o is the speed at which the lead train should be travelling on departure from the platform as the following train approached the platform, for optimum headway to be achieved.

V_o can be determined by finding the speed for which the leading train's acceleration distance and braking distance from / to rest would be equal to the train length plus the safety margin:

$$V_o = \sqrt{\frac{O+L}{\frac{1}{2a_l}+\frac{1}{2b_l}}} \text{ for } V_l \rangle \sqrt{\frac{O+L}{\frac{1}{2a_l}+\frac{1}{2b_l}}} \text{ and } V_o = V_l \text{ otherwise} \qquad \text{Equation F 4-3}$$

If this theoretical scenario is developed to allow for practical headways, as outlined in chapter 8, the



headway becomes as shown in Figure F 4-2.

<u>Figure F 4-2: Relative Braking Headway Compared to Moving block Headway</u>

The plain line headway time is then given by:

$$H_{t\ rb} = \sum_{x_{start}}^{(x_{start}+Pt+Tw+Tp+Tip+bbu_f+Bd_f+O+Tlu+Tlu_e+Tlu_l+Ut+L_l-Bd_l)} \frac{\Delta x}{V_{l\ act}} \qquad \text{Equation F 4-4}$$

It has already been noted that, on approach to a junction, at least a braking distance separation (plus a safety margin) must be maintained between the train and junction, so long as the junction is not detected as locked in the required direction. This means that:

- The headway time equation for a diverging followed by a straight through train will be the same under relative braking separation as under moving block (see equation F 3-1);
- The headway time for consecutively diverging trains will be given by the plain line equation, with allowance for the reduced speed profile (see equation 9-8). This will give a shorter time than that achieved under moving block operation;
- The headway time equation for a straight through train followed by a converging train will be the same under relative braking separation as under moving block (see equation F 3-4);
- The headway time equation for a converging train followed by a straight through train will be given by:

$$H_{t\,rb} = \left(\sum_{x_{start}}^{\left(x_{start}+Pt+Tw+Tp+Tip+bbu_f+Bd_f+O+Tlu+Tlu_e+Tlu_t+Ut+L_t-Bd_t\right)} \frac{\Delta x}{V_{t\ act}} \right) + \left(\sum_{SP_{conv}}^{(FS_{conv})} \frac{\Delta x}{V_{conv.act}(x)} \right) - \left(\sum_{SP_{str}}^{(FS_{conv})} \frac{\Delta x}{V_{str.act}(x)} \right)$$

Equation F 4-5

This will give a shorter time than that achieved under moving block operation;

- The headway time for a train stopping at a single in-line platform will also be given by equation 12-4, with allowance for the reduced speed profile and dwell time. This will give a shorter time than that achieved under moving block operation;
- The headway between consecutive trains stopping in alternate platforms of a two platforms per direction station will be the same as that under moving block operation. That is, in accordance with equation F 3-1, with 'Sg' being replaced by the junction point locking track section, 'Sj' (see section F3.3.1);
- The headway between a stopping train (in an off-line platform) and a non-stopping train travelling straight through (see section F3.3.2) will be given by:

$$H_{t\,rd} = Dwell + \left(\sum_{x_{start}}^{\left(x_{start}+Pt+Tw+Tp+Tip+bbu_f+Bd_f+O+Tlu+Tlu_e+Tlu_t+Ut+L_t-Bd_t\right)} \frac{\Delta x}{V_{t\ act}} \right) + \left(\sum_{SP_{conv}}^{(FS_{conv})} \frac{\Delta x}{V_{conv\ act}(x)} \right) - \left(\sum_{SP_{str}}^{(FS_{conv})} \frac{\Delta x}{V_{str\ act}(x)} \right)$$

Equation F 4-6

This will give a shorter time than that achieved under moving block operation;

- The headway for a terminal station will be determined in the same was as that for moving block operation (see section F3.4).

APPENDIX G – FULL TABLE OF CONTENTS

APPENDIX H – LISTS OF TABLES AND FIGURES

List of Figures:

List of Tables: