Performance Enhancing of Heterogeneous Network through Optimisation and Machine Learning Techniques

Kan Lin

Department of Electronic and Electrical Engineering

The University of Sheffield

This thesis is submitted for the approval of the

Doctor of Philosophy

March 2020

I would like to dedicate this thesis and everything I do to my parents. I would not be who I am today without their love and support.

Acknowledgements

Foremost, I would like to express my sincere thanks to my supervisor Prof. Jie Zhang for the continuous guide and support during my PhD study and related researches. His patience, encouragement and guidance helps me to finish the hold study. He always encouraged me to go deeper in interesting research topics. Besides my supervisor, I would like to thank my second supervisor, Dr. Xiaoli Chu, it gives me great support in my study.

I need to say thanks to my colleagues, Mr. Weijie Qi, Ms. Hui Zheng, Mr. Zimu Zhao, Mr. Baoling Zhang, Mr. Qi Hong, Mr. Haonan Hu, Mr. Bo Ma, Ms. Fan Wu, Mr. Yuan Gao, and Mr. Bozhong Chen, for the troubleshooting and guidance in my simulation, and the time we have in the last four years.

I would also want to send my special thanks to my faculty for all the support they gives during my study.

Finally, I would like to thank my family, for supporting me spiritually throughout the study. It also needs to thank my aunt and uncle and cousin, for supporting me finally in the most crucial time.

Abstract

In the last two decades, by the benefit of advanced wireless technology, growing data service cause the explosive traffic demand, and it brings many new challenges to the network operators. In order to match the growing traffic demand, operators shall deploy new base stations to increase the total cellular network capacity. Meanwhile, a new type of low-power base stations are frequently deployed within the network, providing extra access points to subscribers. However, even the new base station can be operated in low power, the total network energy consumption is still increased proportional to the total number of base station, and considerable network energy consumption will become one of the main issues to the network operators. The way of reducing network energy consumption become crucial, especially in 5G when multiple antennas are deployed within one site. However, the base station cannot be always operated in low power because it will damage the network performance, and power can be only reduced in light-traffic period. Therefore, the way of balancing traffic demand and energy consumption will be come the main investigation direction in this thesis, and how to link the operated power of base station to the current traffic demand is investigated. In this thesis, algorithms and optimisations are utilised to reduce the network energy consumption and improve the network performance.

To reduce the energy consumption in light-traffic period, base stations switch-off strategy is proposed in the first chapter. However, the network performance should be carefully estimated before the switch-off strategy is applied. The NP-hard energy efficiency optimisation problem is summarised, and it proposes the method that some of the base stations can be grouped together due to the limited interference from other Pico cells, reducing the

complexity of the optimisation problem. Meanwhile, simulated annealing is proposed to obtain the optimal base stations combination to achieve optimal energy efficiency. By the optimisation algorithm, it can obtain the optimal PCs combination without scarifying the overall network throughput. The simulation results show that not only the energy consumption can be reduced but also the significant energy efficiency improvement can achieve by the switched-off strategy. The average energy efficiency improvement over thirty simulation is 17.06%.

The second chapter will tackle the issue of how to raise the power of base stations after they are switched off. These base stations shall back to regular power level to prepare the incoming traffic. However, not all base stations shall be back to normal power due to the uneven traffic distribution. By analysing the information within the collected subscriber data, such as moving speed, direction, downlink and time, Naive Bayesian classifier will be utilised to obtain the user movement pattern and predict the future traffic distribution, and the system can know which base station will become the user's destination. The load adaptive power control is utilised to inform the corresponding base stations to increased the transmission power, base stations can prepare for the incoming traffic, avoiding the performance degradation. The simulation results show that the machine learning can accurately predict the destination of the subscriber, achieving average 90.8% accuracy among thirty simulation. The network energy can be saved without damage the network performance after the load adaptive function is applied, the average energy efficiency improvement among three scenarios is 4.3%, the improvement is significant. The significant improvement prove that the proposed machine learning and load adaptive power modification method can help the network reduce the energy consumption.

In the last chapter, it will utilise cell range expansion to tackle the resources issue in cooperative base station in joint transmission, improving downlink performance and tackle the cell-edge problem. Due to the uneven traffic distribution, it will cause the insufficient resources problem in cooperative base station in joint transmission, and the system throughput will be influenced if cooperative base station executes joint transmission in high load. Therefore, the cell range expansion is utilised to solve the problem of unbalanced traffic between base station tier, and flow water algorithm is utilised to tackle the resources distribution issue during the traffic offloading. The simulation shows the NP-hard problem can be sufficiently solved by the flow water algorithm, and the downlink throughput gain can be obtained, it can obtain 26% gain in the M-P scenario, and the gain in P-M scenario is 24%. The result prove that the proposed method can provide significant gain to the subscriber without losing any total network throughput.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

In this chapter, thesis background and motivation will be introduced, and the contribution and thesis structure will be list at the end of this chapter.

1.1 Background

In the last two decades, the advanced wireless technology and fast-developed mobile phone industry bring many new challenges to Cellular Network (CN) operators. In order to match the growing traffic demand, operators need to deploy increasing number of Base Stations (BSs) within the network, but limited space for the BSs deployment has became a main issue to the network operators. Therefore, increasing number of new type low-power BSs will be deployed within the network to increase the total network capacity. Even these BSs have smaller power than the traditional BSs, the total power consumption of the entire network still considerable, and the high energy consumption will significantly reduce the operator's profit. Nowadays, the Energy Efficiency (EE) of CN has already became a critical factor to the network operator. On the other side, when the energy consumption become a fixed value, the EE will be close relative to the Spectrum Efficiency (SE), any improvement or degradation in SE will cause corresponding influence to the EE. Therefore, many researches are hold to determine the method of improving EE via reducing energy consumption or improving SE.

1.1.1 Cellular Network Architecture

The fast-developed manufacturing industry lets mobile phones became an affordable product to the public, and subscriber number is growing exponentially in the last decade. Meanwhile, advanced wireless technology brings tremendous benefits to the way of information exchange, and one significant benefit is the improved downlink speed has already changed the way of people use their mobile phone. Call and message texting will not be the only options in the CN, subscribers can browse the website and watch the video via their mobile terminal, and this changes will let data service becomes the mainstream in contemporary CN. The consequence of this change is the growing traffic demand and backhaul capacity [1], a large amount of data need to be transmitted within the network. With limited spectrum, network operators need to deploy more BSs to match the growing traffic demand.

In Homogeneous Cellular Network (HoNet), most of the BSs are installed on top of the building. The growing traffic demand will force network operator to deploy increasing number of BS to solve the insufficient network capacity issue. However, the Capital Expenditures (CAPEX), including the expenditures of equipment purchasing and limited space hiring, is expensive in urban city because of the limited space is suitable for deploying BS. Therefore, network operators will try to deploy low-power Pico Cell (PC) under Macro Cell (MC) coverage to provide extra access points to its subscribers, and PC deployment not only increase the total network capacity but also improve the subscriber's received signal strength, and these PCs can offloading the traffic from MC [2]. This promising network architecture is called Heterogeneous Network (HetNet), and it becomes the mainstream in future CN [3] [4], and two CN architectures are shown in Figure 1.1.

It can be seen that the energy consumption of the CN will be proportional increased to the number of deployed BSs, and high energy consumption will increase the Operating

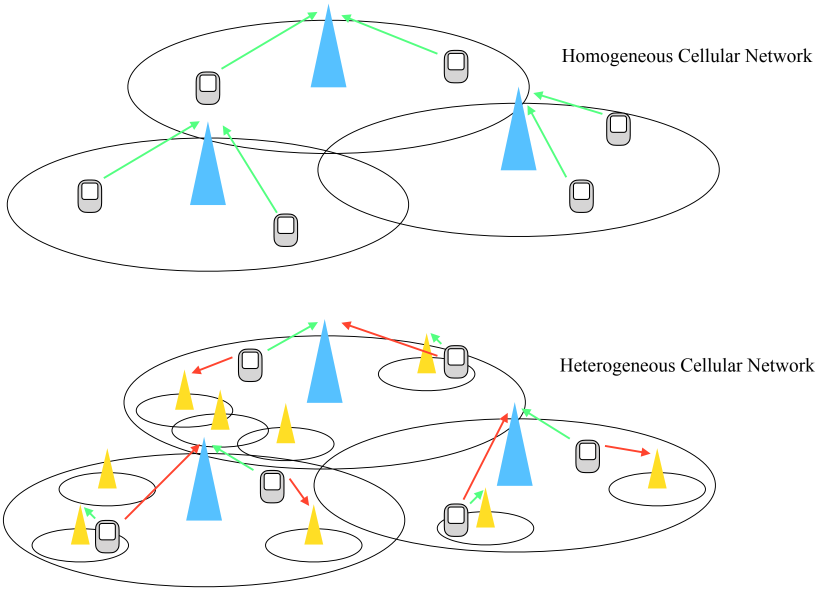


Fig. 1.1 Cellular network architecture.

Expenditure (OPEX), including the cost of repairing, bill for power consumption, engineer recruitment etc., and these costs will reduce the operator's profit. The energy consumption will become the main concern from the network operators point of view. In traditional HoNet, MC should be operated in regular transmission power to guarantee the subscribers can be served anywhere and the power of MC should be a fixed value in most of the cases, any modification of the MC's transmission power might damage or degrade the network performance. Because the PC is deployed within MC coverage in HetNet, subscriber within an MC-PC overlapping area can be served by either one of them, which is shown in Figure 1.1. If the remaining resources of MC are sufficient, the subscriber can be served by the MC, and the light-load PCs can be switched off or operated in low power to reduce the energy consumption of the network. Moreover, based on Figure 1.2 [5] [6], daily traffic is fluctuated from the daytime to the nighttime. The traffic demand between 2 a.m. and 6 a.m. is light, the system will still have sufficient resources to serve the subscribers even some PCs are

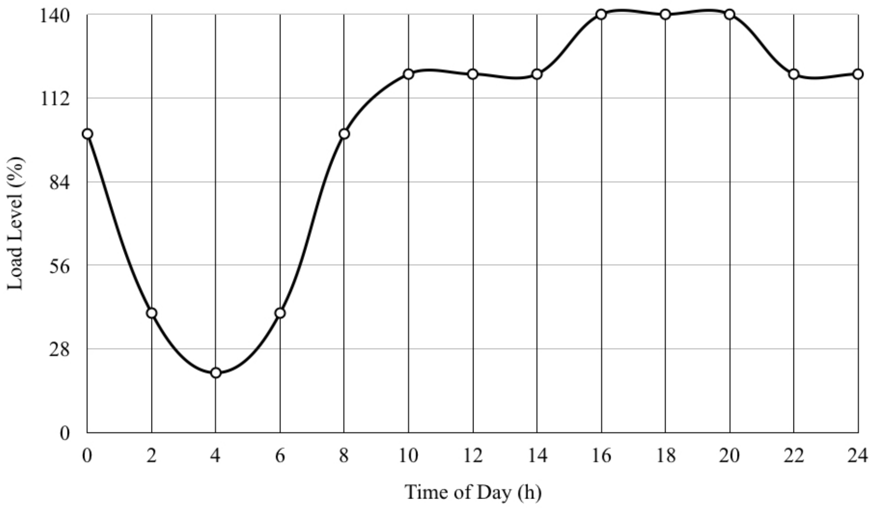


Fig. 1.2 Daily traffic in urban city.

switched off, and the transmission power of PC shall return to regular power level only the traffic demand continues growing.

1.1.2 Self-Organisation Network

Due to the high CAPEX and OPEX, it will reduce the profit of operating the CN, the light-load PCs can be switched off to reduce the OPEX, but the system needs technicians to optimise the parameters to achieve optimal network performance when the network is changed. However, manual testing and configuration will become another part of the OPEX. Because Less manual work will led to lower OPEX and higher profit, network operators are trying to find a new way to achieve the minimum maintenance with guaranteed network performance, and the self-organisation idea attracts both researchers and operators interest. Nowadays, Institute of Electrical and Electronics Engineers (IEEE), 3rd Generation Partnership Project (3GPP) and some other organisations have already defined the Self-organisation

Network (SoN) as a network that it can achieve many functionalities without or with low manual operations, and SoN is considered as one of the key features in the future CN [7] [8]. The system can automatically search the environment and configure the parameters to achieve optimal network performance in SoN, and there are three categories are included in SoN, such as self-configuration, self-optimisation and self-healing. The aim of self-organisation network can be included to achieve the following aims:

- To reduce the CAPEX and OPEX of operating the network, and the system can automatically self-maintain with low manual maintenance.

- It can achieve an optimal network performance by adjusting the parameters, optimising coverage, throughput, energy efficiency etc.

- To search the surrounding environment of BS and detect any new BSs are added around it, it can configure parameters to adapt the new environment.

In the future, increasing number of PCs are deployed to match the growing traffic demand. The air interference will become more complicated because the interference will come from different BS tier. Moreover, due to the dynamic traffic in urban city, the traffic demand will be changed in each hour, and some parameters need to be configured to adapt the new traffic demand and complicated wireless interface. In order to achieve optimal network performance, some parameters need to be configured when new BSs are added in the network, and traffic, allocated resources of BS and other information can be shared within BSs in SoN. If a new PC is detected, the new PC's information shall be shared within the CN. Therefore, it needs plenty of technicians involve the testing and configuration, and it will become a challenge for the network operators to obtain an optimal network performance with low manual operations. Testing and configuration can be significantly reduced by the benefit of SoN because some functionalities are added into BSs that allow the BSs to efficiently configure parameters without manual works.

1.1.3 Energy Consumption of Cellular Network

The energy consumption of CN can be mainly divided into three parts, including mobile terminal, BS and the core network. The power consumption in each part will be described in detailed below [9] [10].

Mobile Terminal

Due to the tremendous subscriber number, the power consumption of mobile terminal is considerable, and the energy is mainly used in the application, signal transmission, display etc. However, the energy consumption of mobile terminal will not reduce the cost of operating CN, operators play fewer interests in reducing the energy consumption of mobile terminal.

Base Station

It is worth to notice that almost 80% of the CN energy is consumed by BS, and reducing the energy consumption of BS can significantly reduce the energy consumption of the network [9]. The energy consumption of BS can be generally divided into four parts, air conditioning, signal processing, power supply and signal transmission, power consumption of each part is shown in Figure 1.3 [11]. In these four parts, the power supply provides a basic incoming power source, and it keeps other BS's functionalities work normally. The air conditioning is another essential part in BS, the reason for installing a cooling system is to reduce the heat from the machine, and it can significantly reduce the chance of fire and burning. Therefore, the power of cooling system should be guaranteed at any timing. The received signal of BS will be processed before transmitted to other BSs, and the signal processing should be operated in normal condition when BS is switched on. It can be seen that air conditioning, signal processing and power supply are the fundamental parts of BS, and it is used to maintain the regular functionalities. Therefore, the energy consumption in these parts cannot be reduced because any energy reduction could cause a damage or influence on BS operation. The power for signal transmission will become the main focus in this thesis, some BSs in low-load status can be switched off to reduce the energy consumption of CN, especially in light traffic period [6].

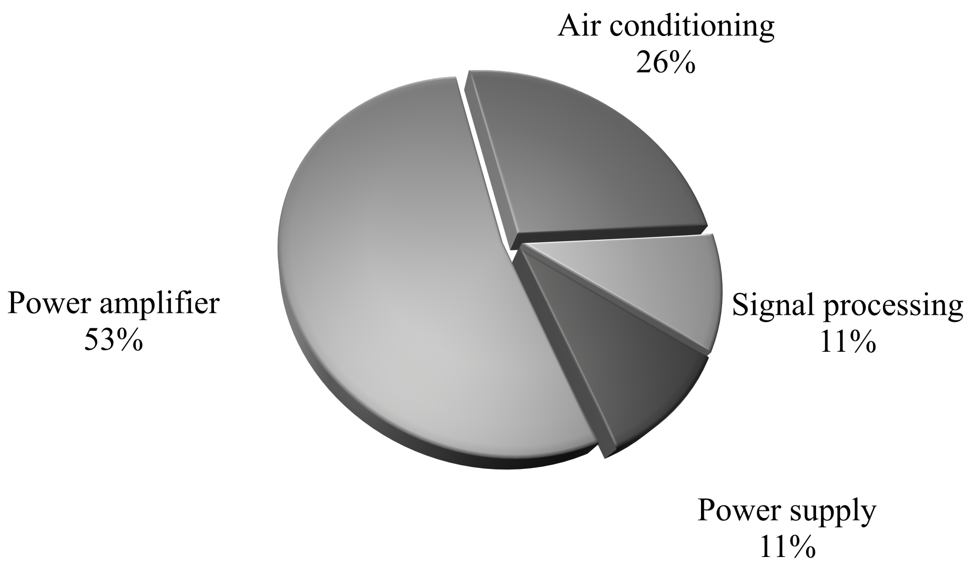


Fig. 1.3 Power consumption of base station.

Core Network

As the information exchange centre, the power supply of core network should be guaranteed at any timing. The failure power supply will cause severe loss to the operators, and energy consumption of this part is not included in this thesis.

1.2 Motivation

Many efforts are taken to build a green CN and improve the EE. The EE of CN is closely related to SE and total energy consumption of the network, and these two factors catch network operators interest. Currently, the high-density BS deployment will become the trend of CN deployment, especially in 5G. However, it will increase the OPEX due to the expensive bill in energy consumption. To reduce the network energy consumption, because most of the energy is consumed by BS, reducing the BS power consumption can significantly reduce the power consumption of CN. In another aspect, SE of CN is another way to improve the EE when the power consumption of the CN is fixed. Based on these two factors, the following two points can be further investigated, which is power consumption reduction and SE enhancement.

1.2.1 Power Consumption Reduction

The flowing daily traffic gives a chance to reduce the energy consumption of the network. In high-density BS deployment, the objective of this type of network deployment is to match the growing traffic demand due to the limitation in MC deployment. However, daily traffic demand fluctuates every hour, and only small number of BSs can sufficiently serve the existing subscribers during the light-traffic period. Therefore, the remaining BSs can be operated in low-power mode or completely switched off to reduce the energy consumption, but the switched-off BSs might cause the network performance degradation. Therefore, the methods for balancing the energy consumption and traffic demand shall be proposed to avoid the impact to the network. Meanwhile, the CN needs a method to let these switch-off BSs back to regular power level when the traffic demand is increasing. However, the traffic is not evenly distributed, it is not necessary to let all BSs operate in full power. It can be seen that the traffic prediction is needed to inform the switched-off BSs wake up to prepare for the incoming traffic in the right timing, reducing the energy consumption. Because the system can obtain a high accuracy traffic prediction, the power of BSs can be increased proportional to the load of BSs. Therefore, reliable traffic prediction method is needed for investigation.

1.2.2 Spectrum Efficiency Enhancement

The above part mention the energy saving can help the improvement of EE. However, reducing the transmission power of BS during heavy-traffic period might not be a wise decision because it will cause many subscribers will out of service due to a smaller coverage than BS with normal transmission power. Meanwhile, the resources of BS cannot be fully utilised. To improve the EE in high-traffic period, the other way is to improve the SE, and Joint Transmission (JT) is proposed to tackle the unsatisfied downlink performance problem, typically the cell-edge users performance. However, if the cooperative BS is suffered from traffic congestion, the network throughput will be influenced by the JT. Therefore, methods need to be investigated to solve the traffic congestion problem in cooperative BS, and JT can be executed without sacrificing network throughput.

1.2.3 Hypothesis

Thus, this thesis is mainly focused on EE from light-traffic period to mid-traffic period in order to provide an optimal network performance with low energy consumption.

The research questions in chapter three can be summarised as:

• The possibility of switching light-load PCs to reduce the total energy consumption without reducing the subscriber's downlink throughput.

• The EE can be improved after the a certain number of PCs are switched off, the proposed SA could determine the optimal PCs combination to maximise the EE rather than simply turn off the original light-load PCs.

The research questions in chapter four can be summarised as:

• The subscriber data can be analysed by machine learning, and it can build the connection between subscriber and BSs.

• The NBC can utilise the data analysis result from the subscriber data to prediction the traffic distribution.

• The PC's transmission power can be changed according to the predicted load of PC, reducing the total network energy consumption and obtaining better EE.

• The EE improvement will not be significant when the load of the network traffic reaches medium or high.

The research questions in chapter five can be summarised as:

• The load of cooperative BS will cause impact on the performance of the JT, selecting cooperative BS shall not only consider the signal level of BS.

• Flow water algorithm is used to tackle the traffic distribution issue. After the algorithm is applied, each BS can have almost the same load within the network.

• Cell Range Expansion (CRE) is applied to heavy-load BS to transfer the load to nearby light-load BSs, and the JT can be executed without reducing the total network performance.

1.3 Contributions of the Thesis

As a result, the contribution of the thesis can be summarised as follows:

• To reduce the energy consumption during light-traffic period, and techniques are applied to optimise the EE of network, and it includes: 1) BSs switched-off strategy is applied to reduce the energy consumption of the network 2) The EE is formulated as an optimisation problem, and sub-optimisation problem is formulated to reduce the computation complexity. 3) The simulation proves that the best PCs combination is different with switching off original light-load BSs.

• To predict the traffic distribution, and the prediction result will be utilised to prepare the incoming traffic and minimise the energy consumption of network, and it includes: 1) Subscriber's data is collected and analysed, Naive Bayesian Classifer (NBC) is utilised to determine the characteristic between two PCs. The NBC will give the traffic prediction based on the collected subscribers' data. 2) According to the traffic prediction, the system will increase the transmission power of PC to prepare the incoming traffic. 3) The energy consumption and OPEX can be reduced, improving the EE.

- To further improve the spectrum efficiency via JT. Techniques is applied to balance the traffic between two BS tiers, and heavy-load cooperative BS can execute the JT without sacrificing the network throughput, and it includes: 1) JT downlink performance with different cooperative BS selection. 2) Flow water algorithm is utilised to tackle the optimisation problem. 3) CRE is applied to offload the traffic from high-load cooperative BS to low-tier BS, and the JT can be executed without sacrificing the network throughput.

1.3.1 Publication

K. Lin and J. Zhang, "Energy efficiency enhancement via power modification with simulated annealing cell selection," *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC),* Las Vegas, NV, 2018, pp. 526-531.

1.4 Structure of the Thesis

The thesis is organised as follows:

Chapter 2: Literature Review

This chapter introduces the background knowledge of EE, Machine Learning (ML), traffic prediction and JT, it also provides the previous works and represents the potential research gap in the related area.

Chapter 3: Energy Saving with Optimal Cells Association in Heterogeneous Network

Due to the light-traffic period occurs at the early morning in urban city, the unnecessary operating PC in CN will increase the power consumption. Some light-load PCs can be switched off to reduce the energy consumption without degrading network performance. After some light-load PCs are switched off, subscribers can be served by neighbour PC, but some of them might receive weak signal from new serving BS because the closest BS is switched off. Therefore, the system should choose the best PCs combination that it can provide optimal network performance. Simulated Annealing (SA) algorithm is utilised to determine the best PCs combination.

Chapter 4: Energy Efficiency Enhancement via Naive Bayesian Classifier Traffic Prediction

Traffic will raise to medium or high level when time is approaching working hour, the system shall switch on PCs to prepare the incoming traffic. Moreover, commuters have a fixed moving pattern when they appear around their working places, and subscriber information will become a valuable data for the system to learn and predict their further movement, NBC will become a ideal tool to achieve this objective. The system can prepare the incoming traffic and minimise the energy consumption of network via prediction result.

Chapter 5: Spectrum Efficiency Improvement via Joint Transmission and Cell Range Expansion

JT is proposed to minimise the interference from adjacent BS, improving the cell-edge performance. However, this technic requires more than one BSs to transmit signal to one subscriber. The cooperative BS might face the problem of resources shortage because the uneven traffic distribution, and JT might reduce the network throughput if cooperative BS is in high-load. CRE is applied to offload the traffic from MC, and JT can be executed after the resources shortage problem is tackled .

Chapter 6: Conclusions and Future Work

The conclusion and future work is summarised.

Chapter 2

Literature Review

Topics in the current literature related to the thesis are reviewed in this chapter. In this chapter, it includes the EE and BS switch-off strategy in the first section. The ML and traffic prediction details are presented in sections two and three, and JT is introduced in section four. Related works will also be discussed at the end of each part.

2.1 Energy Efficiency in Cellular Network

2.1.1 Definition

Network operators are mainly concerned with core network features (coverage, service, etc). However, the definition of EE is presented differently in different work owing to different research goals. These goals include the total throughput of multiple BSs divided by total energy consumption in [12], the power consumption per covered area in [13], and the power consumption per capacity in [14]. Furthermore, these definitions have two main areas of focus: the total energy consumption of the network, and the benefit that operators can be obtained when energy is consumed.

In the following chapters, the definition of EE will be represented by Equation 2.1, which was calculated as bits-per-Joule [15] - [17]. T represents the total network throughput and P

represents the total power consumption. Under full coverage conditions, network operators will focus on the total volume of data that can be transmitted when all BSs are operating at full power. Thus, EE can give network operators a better understanding of network performance.

$$E_n = \frac{T}{P} \tag{2.1}$$

According to the above definition, the EE of a network can be enhanced by either improving network throughput or reducing power consumption, with the optimal situation being one where the system can achieve both.

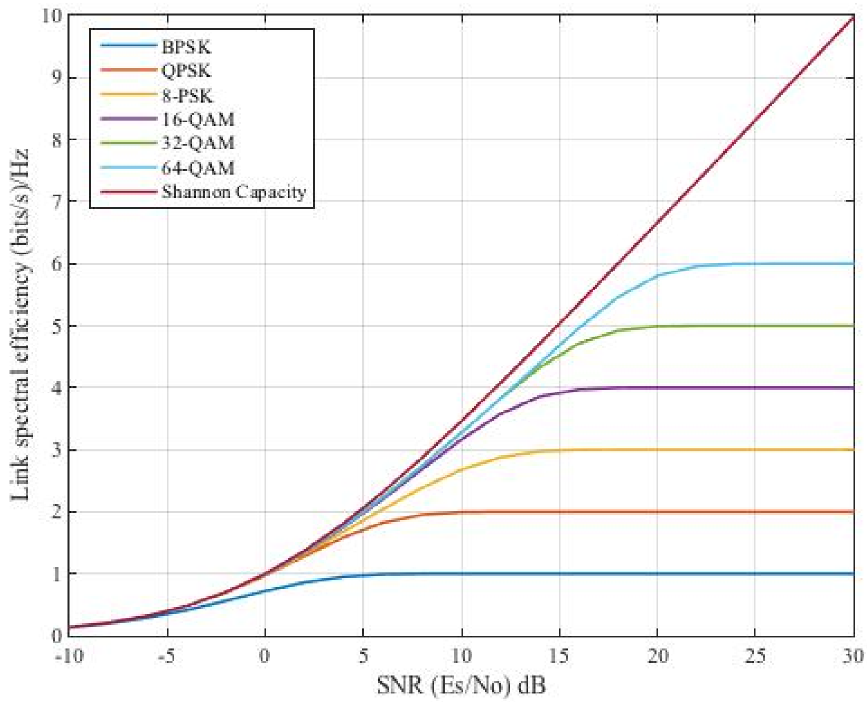


Fig. 2.1 Spectrum efficiency.

The network throughput is equal to the sum of BS throughputs, which are directly related to spectrum efficiency and BS bandwidth. According to the aforementioned equation, maximising spectrum efficiency can significantly increase BS throughput when the bandwidth is a fixed value, which is shown in Figure 2.1. It is clear that optimal spectrum efficiency will also improve EE. The BS throughput is calculated as the sum of the subscriber throughputs;

however, subscribers are randomly distributed within the network. Thus, the distance between BS and subscribers is varied. In addition, the signal subscribers receive will weaken as their distance from the BS increases [18] [19]. Moreover, spectrum efficiency in outdoor environments is further complicated by the fact that obstacles between subscribers and the BS will absorb energy. Thus, improving spectrum efficiency is a key challenge for network operators.

The total network power consumption comprises four parts, which were introduced in the previous chapter. The BS power consumption is considerable but can be reduced when traffic demand is light [10]. However, there is a trade-off between power reduction and coverage. The BS requires a certain level of power to provide full coverage to subscribers, and transmission power will significantly influence the subscribers's received signal strength [18]. If the signal strength and allocated resources cannot reach the minimum required values, it will degrade the quality of service (QoS). Hence, reducing BS power consumption should guarantee that subscriber QoS is not degraded. Many pieces of focus on different approachs to improving EE (new network architecture, improving power amplifier efficiency, improving radio interfaces, etc) [17]. Concerning network architecture, research has shown that HetNet can effectively improve EE [14] [20]. In addition, PCs operate at low power levels, meaning the coverage for a PC will be significantly smaller than that of an MC. The benefit of high-density PC deployment is that the path loss can be significantly reduced because of the shorter distance, thereby ensuring subscribers can receive an optimal BS signal. Thus, more data can be transmitted under a fixed bandwidth; however, it is clear whether high-density PCs deployment can increase the total energy consumption (owing to traffic fluctuation) [14].

Compared with MCs, PCs require a small amount of power for signal transmission; the power for signal processing becomes the major part of BS energy consumption. In [14], the downlink performance is significantly improved by virtue of the advanced radio interface, which means that the network can obtain a higher spectrum efficiency than before

[20]. However, high spectrum efficiency transmission techniques also require complex computation and greater power for signal processing. Thus, power consumption during signal processing becomes another issue in reducing BS energy consumption.

2.1.2 Base Stations Switch on and off Strategy

Many authors have proposed that BSs can be switched on and off to reduce energy consumption. Hence, two BS modes (active mode and sleep mode) have been proposed and applied [21] - [24]. When a network faces a high traffic demand, the BS will enter active mode to serve the subscribers; sleep mode will be activated in low traffic periods.

Active Mode Active mode represents a BS under normal operating conditions (including the cooling system, signal processing, power amplifier etc). Signals are transmitted at normal power levels, and all subscribers may access the network without exceeding the BS's capacity.

Sleep Mode The main characteristic of sleep mode is that some hardware components are operating at low power or are completely deactivated, which has been proposed in [24]. The sleeping components in PC mainly depend on the specific sleeping mode type or algorithm.

Typically, three scenarios are taken as an example of the switch on and off strategy [25] [26]. The scenarios are almost under the same condition that the network traffic is fluctuated by the active subscribers.

In general, three scenarios are taken as examples of this strategy [25] [26]. These scenarios generally occur under two conditions where network traffic is fluctuated by active subscribers:

• In intense traffic areas, multiple BSs cover the same geographical area, and one BS can be switched to conserve energy when traffic demand is light.

• Two-tier BS deployment; MCs provide coverage in a large area, and PCs will provide capacity and access points in specific locations.

• In the areas where a high-speed BS is totally covered by a traditional BS, the high-speed BS can be switched off when there is no demand for high-speed data.

However, the system should avoid frequently switching BSs on and off [27]. Moreover, the load condition of BSs should be considered, as part of the traffic will be transferred to neighbouring BSs when they are switched off. If one BS has already suffered from traffic congestion, the transferred traffic will further exacerbate this problem. In addition, the system must consider many factors before making on which BS to deactivate. Thus, this strategy has much potential for future study.

2.1.3 Literature Review on Base Stations Switch on and off Strategy

In [21], a genetic algorithm-based cell deactivation strategy for improving network EE is proposed. In this study, the problem is formulated as a binary integer linear programming (ILP) problem. Considering BS numbers as a sequence, this approach attempts to deactivate the BS with a sequence number. Using the idea of the genetic algorithm, it changes part of the sequence number and generates offspring, which converge after serval iterations. However, considerable time is required to compute the solution. Meanwhile, another study uses the same algorithm to get the best switched-off strategy [28]. This study considers the load of BS and the neighbour BSs, but the study did not take the consideration of the transmission distance between BS and subscribers. If some of the nearby BSs are switched off, subscribers shall connect to the remote BS, and subscribers will need more resources to obtain the same downlink speed.

In [27], the authors aim to avoid frequent switching between active mode and sleep mode. A dynamic programming algorithm is used to solve the time-correlated problem, which is formulated as a dynamic programming problem that it contains elements of the state, action, state transition, and previous state cost. The result guarantees the system can obtain satisfied call block probability, and the study proves that the active BSs can match network traffic in the temporal and spatial domains. By adjusting parameters, energy savings and energy cost of switching between two modes can be balanced.

The study in [29] consider the resources of the BSs and the required downlink of subscribers, and the study summarised the problem as a Integer Linear Programming (ILP). The study shows that some BSs can be switched off to save the energy, and it can do the optimisation based on the traffic of the network. However, the study is lack of the consideration of total resources usage of the network. The study do the optimisation based on all subscribers can be fulfilled with their requirement, the same issue can be found in [30].

The BSs zooming method was proposed to tackle coverage problems where some BSs are switched off in CN [31]. In essence, BS zooming can solve uneven traffic distribution, offloading traffic from a congested BS to a low-load neighbour BSs. Moreover, the system can deactivate some low-traffic BSs to reduce energy consumption during light traffic periods, while original subscribers can be served by neighbouring BSs. The extensive cooperation between BSs, enables the system to deactivate any BS without QoS degradation.

In [32], an optimal on-off pattern is proposed, which considers the interference from neighbouring BSs. In this method, mixed-integer convex optimisation is used to determine the proportion of BSs that must be obtained to serve the subscribers. Following this, Markov analysis is applied to the system (two independent Markov processes are presented in the paper). Finally, a heuristic cooperative control method is applied in the dynamic scenario to improve EE.

A cross-layer optimisation method is proposed in [33] to reduce energy consumption; PCs are deployed within MC coverage. Once some BSs match the criteria, the system will switch off to save energy. Meanwhile, the system must decide user associations when the original serving BS is deactivated. A similar method is proposed in [34], where the low load BSs will be deactivated and original serving BS is replaced by an MC. By comparing the energy conservation and transferring traffic, the system can determine the best BS deactivation strategy.

Note that BSs can be deactivated where the current load is not heavy and where QoS will not be degraded when the MC serves the remaining subscribers [35]. The continuous time Markov decision process is applied to represent the load of every BS.

In conclusion, this strategy becomes challenging because many other variables(interference, received signal quality, etc.) will influence EE. Moreover, when algorithm is applied in a real network, computation becomes difficult owing to large numbers of BSs and subscribers. Thus, further study is needed to optimise this.

2.2 Machine Learning

Machine learning (ML) was introduced in 1959, and it can be defined as a computational method that can make predictions or improvements based on input data. The practical applications of ML are to predict unforeseen items or data and to learn and build efficient algorithms for these predictions. Examples of input data are product information or electronic data taken from Internet; however, ML requires considerable amounts of data to provide optimal prediction accuracy [36] - [38]. In fraud detection, recognition, and document classification, ML is a promising tool when large amounts of data are available. After much development, ML can be mainly classified into two categories, supervised and unsupervised learning, which are briefly described below [38].

Supervised Learning In supervised learning, a set of training data is provided with a corresponding set of labels, predictions are made based on this input data. By analysing the input data, the functions or the relationships that will be utilised for further prediction are constructed. A typical example is junk e-mail filtering, where previous data is used to build a probability relationship between words and junk e-mail. Thus, new e-mail will be defined as junk depending on the words that appear within [39].

Unsupervised Learning In unsupervised learning, data are provided without labels and predictions are made based on previous unlabelled data. This means that structures and relationships must be independently identified or built. Since the data is provided without labels, evaluating the accuracy of the structures is difficult [38]. Clustering, wherein data belong to the cluster with the smallest mean, is an example of an unsupervised learning problem.

2.2.1 Probabilistic Classifiers

Probabilistic classifiers provide output with the probability that new data belongs to a specific class. Therefore, it assumes for a group from class $\{1,2,3,....c\} \in C$ and a set of collected data x, the probability of new data for class c can be represented by $P(c \mid x_{new})$. The probability should also follow the constraints in [40], which is shown in Equations 2.2 and 2.3.

$$0 \leq P(c \mid x_{new}) \leq 1 \tag{2.2}$$

$$\sum_{c=1}^{C} P(c \mid x_{new}) = 1 \tag{2.3}$$

Thus, predictions are made with this probabilistic classifier when the system collects the new data. Some of the advantages of using probabilistic classifiers are provided in [41].

Prediction Reject and Preparation Probability is useful in some applications because it provides a certain confidence in the output. In cases of assigning a data into a suitable class, these predictions seem uninformative if they only provide a class number. However, probabilistic classifiers can provide probability, where $P(c_1 \mid x_{new}) = 0.5$, $P(c_2 \mid x_{new}) = 0.4$

etc. In this case, the prediction becomes useful when the probability does not match the minimum value for accepting this classification, and it can prepare for the second highest probability prediction if the highest one is incorrect.

Features Analysis In the classification, the features of the data in x, which determine the degree of probability for the class c are analysed. In some applications, the most reliable features are preferred. Some features might not be sufficiently related to the class. Hence, more than one feature in a given classification can be combined. It is assumed that features x_1 and x_2 are independent, and it obtain the probability of class c, where $P(x_1, x_1 \mid c) = P(x_1 \mid c)P(x_2 \mid c)$.

Compensation of Class Imbalance In cases where some outputs are rare in classification, the classifier will give the most common prediction regardless of input data types. However, the training data can be balanced to avoid this problem. It is assumed that $P'(c)$ and $P'(c \mid x)$ after balancing with original data of $P(x \mid c)$. Therefore, it will have the following relationship:

$$P'(c \mid x) \propto P(x \mid c)P'(c) \tag{2.4}$$

Therefore, a posterior can be changed into a likelihood, which is represented as follows:

$$P(c \mid x) \propto P(x \mid c)P(c) \propto \frac{P'(c \mid x)}{P'(c)}P(c) \tag{2.5}$$

2.2.2 Naive Bayesian Classifier

Bayes Theorem Bayes theorem is a sophisticated probability theorem that is applicable in many areas. The theorem represents the probability of an event, which is based on other conditions related to the event [42] [43]. Following the previous notations, class c and input

x can obtain the conditional probability of $P(c \mid x)$ and $P(x \mid c)$. The Bayes' theorem can be represented as follows:

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} \tag{2.6}$$

Naive Bayesian Classifier The Naive Bayesian Classifier (NBC) is a method of supervised learning constructed using the Bayes theorem and probabilistic classifiers. According to Equation 2.6, the numerator $P(x \mid c)$ can be obtained by observation, and the denominator is a constant given the event of c. Using the maximum a posteriori probability (MAP) estimate, the input data is labelled with the maximum value of $\prod_{n=1}^{N} P(x_n \mid c)$ [41] and [44]. The detail and the derivation of NBC will be shown in Chapter 4.

The NBC presents several challenges. First, feature selection significantly influences the prediction results. Since probability calculation is based on feature values, if the chosen features are not related to the class, the classifier will always return an incorrect prediction; the classifier cannot construct or determine a relationship between features and classes. In addition, the probability distribution of each feature could pose a challenge as to whether a particular type of probability distribution is suitable for a given feature; inappropriate probability distribution functions may slightly reduce prediction accuracy.

2.2.3 Literature Review in Naive Bayesian Classifier Prediction

Bayes Theorem is also applied in location prediction in CN by combining neural networks, the products of which are called Bayesian neural networks. This model is used for movement prediction [45]. In detail, the system can access historical data during the learning process and provide predictions for preparing resources for subscribers. Simulation results show this method provides superior prediction accuracy when compared with other standard neural networks.

NBC is frequently used in diagnosis, and is applied for automated diagnoses in cellular networks [46]. This is accomplished by modelling symptoms and conditions as input data, and defining the output as a cause. With the highest probability of cause, the system can determine what causes dropped call in CN. Compared with expert diagnosis, simulation results showed that the classification accuracy reached 71%, and could reach 91% where the two highest classification probabilities are considered.

In [47], logistic regression and an NBC are applied to predict traffic. Moreover, both prediction algorithms can provide offline calculations and fast real-time calculations. First, the model records subscribers' historical data, and an NBC is then used to predict traffic. With the benefit of traffic prediction, the system can prepare its resources for subscribers. Owing to this preparation, QoS can be guaranteed and subscriber satisfaction can be improved.

After collecting the data of estimated time of arrival, transmission, and the distance between receiver and transmitter, an NBC is applied to predict the number of active nodes [48]. Next, the accuracy of this prediction is tested with different feature combinations. Results show that accuracy can be improved if the number of involved features is increased. In addition, the prediction performance is slightly improved after the prior distribution is modified over the appearing times in the class.

2.3 Traffic Prediction in Cellular Network

Prediction is a challenging endeavour in wireless communications, as traffic flow is always linked to subscriber movement. Moreover, prediction mainly relies on object history or a long period of study on a specific object to summarise a pattern. In the following section, we present some of the related works on traffic prediction in CN.

2.3.1 Literature Review in Cellular Network Traffic Prediction

The traditional method of traffic prediction requires predicting subscriber locations. In [49], four past subscriber location points with coordinate information are recorded. By fitting these points into a linear equation, new linear equation that is close to these points is derived and used to predict the next three location points. This method uses the last subscriber location and predicted three locations to draw an elliptical location area that this subscriber might visit in the future. This real-time calculation is simple; however, this method is based on location information, meaning any error will significantly influence the prediction accuracy.

The Markov model is used for traffic prediction in [50] [51]. In detail, this model considers each BS as a state in a Markov chain and defines the traffic transaction probability of each state as an element in the Markov chain. The transition probability matrix represents the probability of one BS transferring the traffic to another BS in several steps. This prediction method can reduce computation complexity, but cannot provide the subscribers movements in detail. A discrete time Markov chain is proposed for the traffic prediction in [52]. Here, the probability transition matrix is constructed using the probability of transitioning from one BS to all remaining BSs. Thus, accurate predictions can be provided after multiple probability transition matrices are calculated. This method does not require any learning processes, and the system can update after a prediction is made. Moreover, the system can also update whenever a handover occurs within the CN. A similar method is proposed in [53]. This method applies normalised probability distribution and logarithmic summation to tackle the calculation issues in the hidden Markov model.

Improvements to the previous method are proposed in [54]. This approach defines a different class of user mobility, which contains a respective probability transition matrix. Next, an expectation maximisation algorithm is applied to find the local maximum value of the missing probability in each prediction. Thus, the model can obtain the probability of one user belonging to one mobility model and the probability of this user moving from one

BS to the another within this model. Finally, although this method increases computation complexity, prediction accuracy is significantly improved.

At the early stage, many researchers realise that the subscriber distribution has such patterns in temporal and spatial [55] and [56]. The advanced method is proposed in [57] that the traffic prediction is made based on the collected subscriber's data. The study discover that the data has a certain pattern that it will aggregate not only in temporal aspect but also in the spatial aspect. Even the traffic prediction is made based on the real subscriber's data, the study still using the traditional method to analyse the data, and the traffic prediction result is not utilised in this study.

In [58], the use of a support vector machine as a classifier to predict the next BSs is proposed. This method records set Channel State Information (CSI) when the subscriber enters and leave a BS and combines the current BS location, CSI sequence, and handover history as input data. Following this, subsequent BS indices are defined as outputs. By learning the input and output data, the classifier can generate a pattern for traffic prediction. In real time traffic prediction, it will provide feedback to improve prediction accuracy and avoid the influences of environmental changes. Thus, this dynamic classifier modification can significantly improve prediction accuracy.

In [59], traffic prediction is modified and divided into three phases. In the first phase, activity types are defined and the probability of these activities (or traffic types) are calculated under existing geographical and temporal constraints. Through weekly user behaviour analysis, this method can construct relationships between time, location, and traffic type and build probability relationships between two activities. In the second phase, the relationships between activities are constructed based on use observation and time-related transition probability matrix is then defined. This matrix can be utilised to predict a user's next location in the final phase.

A recurrent neural network-based mobility prediction method is proposed in [60]. Long short-term memory (LTSM)is applied to tackle disadvantages in network by unbalancing the considerations between the previous input and output. Thus, previous data, current data, and cell information are all put into consideration, and no previous data are wasted. However, the training and calculation time of this method are considerable; further refinement is required to address this challenge.

In the recent study, the ML is proposed to predict the traffic in CN [61]. The research utilise historical, spatial and temporal data to prediction the traffic distribution and utilise the prediction result to reduce network energy consumption. The study mainly focus on the location of the subscriber and did not consider the detail of the subscriber because the traffic demand within CN will not only be influenced by the location of subscriber but also the individual traffic demand.

To conclude, there are several phases of traffic prediction in CN: subscriber collection and analysis, learning, and prediction and updating [62]. Data collection is dependent on the prediction methods, and these features are all closely related to subscriber movement and behaviour. Learning complexity is dependent on the prediction algorithm, and most algorithms finish the learning process offline. In the prediction and update phase, all prediction methods provide their predictions immediately, but some methods do not support instant updates. This causes a reduction in accuracy when subscriber behaviours or environments change.

2.4 Coordinate Multi Points

In the early stages of Coordinated Multipoint Points (CoMP), research was conducted to discover the potential of this technique. This research showed that the average spectrum efficiency and cell-edge performance can be improved by CoMP [63]. This technique was proposed to tackle issues related to cell-edge performance, high data rates, and increasing system throughput in CN [64] [65]. 3GPP defines four scenarios for deploying CoMP[64], which are as follows:

- HoNet with intra-cell CoMP; different sectors within a BS will cooperatively participate.

• HoNet with high transmission power Remote Radio Heads (RRHs); the RRHs are connected to one BS via optical fibre, and cooperation can be established between different RRHs.

• HetNet with low power RRHs within the macrocell coverage; the transmission and reception points created by the RRHs have cell IDs different from those of the macrocell

• HetNet with low power RRHs within the macrocell coverage; the transmission or reception points created by the RRHs have the same cell IDs as the macro cell

In downlink CoMP, these are categorised as joint processing (JP), coordinated scheduling/beamforming (CS/CB). In addition, JP can be categorised as JT and Dynamic Point Selection (DPS) [64] [66] [67]. In this thesis, only JT is analysed. Therefore, the remainder of this paper will describe JT in detail, including reasons for proposing JT, a description of JT, and the related works and discussions.

2.4.1 Joint Transmission

Interference in Heterogeneous Network In HoNet, interference in overlapping areas is an issue that reduces overall network performance. However, there are more interference sources than those from HoNet, owing to high-density PC deployment. Moreover, in order to fully utilise the spectrum and increase network capacity, the best solution where is both MCs and PCs are using the same spectrum to serve more subscribers. However, using the same spectrum to transmit the signal to will cause interference for the subscribers of other BSs. This scenario is shown in Figure 2.2 The blue solid arrow represents the signal from the serving BS, and the remaining red dashed arrow represents interference from other BSs.

It is assumed that the MCs interference group is $\{1,2,3,...m\} \in m$ and PCs interference group is $\{1,2,3,...p\} \in p$. Equation 2.7 shows the relationship between received signal strength and interference.

$$SINR = \frac{S}{\sum_{m=1}^{m} I_m + \sum_{p=1}^{p} I_p + \sigma^2} \tag{2.7}$$

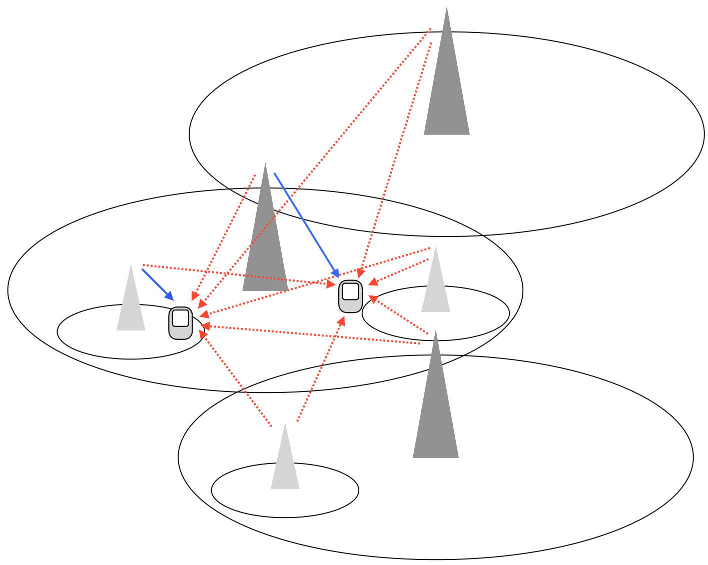


Fig. 2.2 Interference in heterogeneous network.

where S represents the received signal strength of BS, which is dependent on the serving BS. It can be seen that the number of interference sources is proportional to the number of BSs. In HetNet, high-density BS deployment will cause severe interference for other BSs. Some techniques are proposed to mitigate the inter-cell interference or cross-layer interference to address this problem. These include fractional frequency reuse [65] [68], soft frequency reuse [69], enhanced inter-cell interference coordination [70] [71], and CoMP.

Joint Transmission JT is defined as a downlink CoMP that implies dynamic coordination among multiple geographically separated transmission points [64]. Moreover, a JT scheme in downlink is considered to be a potential solution to improve spectrum efficiency and fulfil customer requirements; specifically, improving cell-edge performance, which is shown in Figure 2.3. In normal scenario, when subscriber receives the signal from serving BS, which is shown as blue arrow. However, it will also receive multiple signals from the remaining BS, which is shown as red dash arrow. It can be seen that the subscriber on serving BS edge will receive almost the same signal from interfering BS, and it will bring down the SINR. Subscriber will be more difficult to obtain the required downlink speed. In JT scenario, the interfering signal will become useful signal and help the subscriber on cell edge to obtain optimal downlink speed. Signals from serving BS and cooperative BS will both make contribution in downlink.

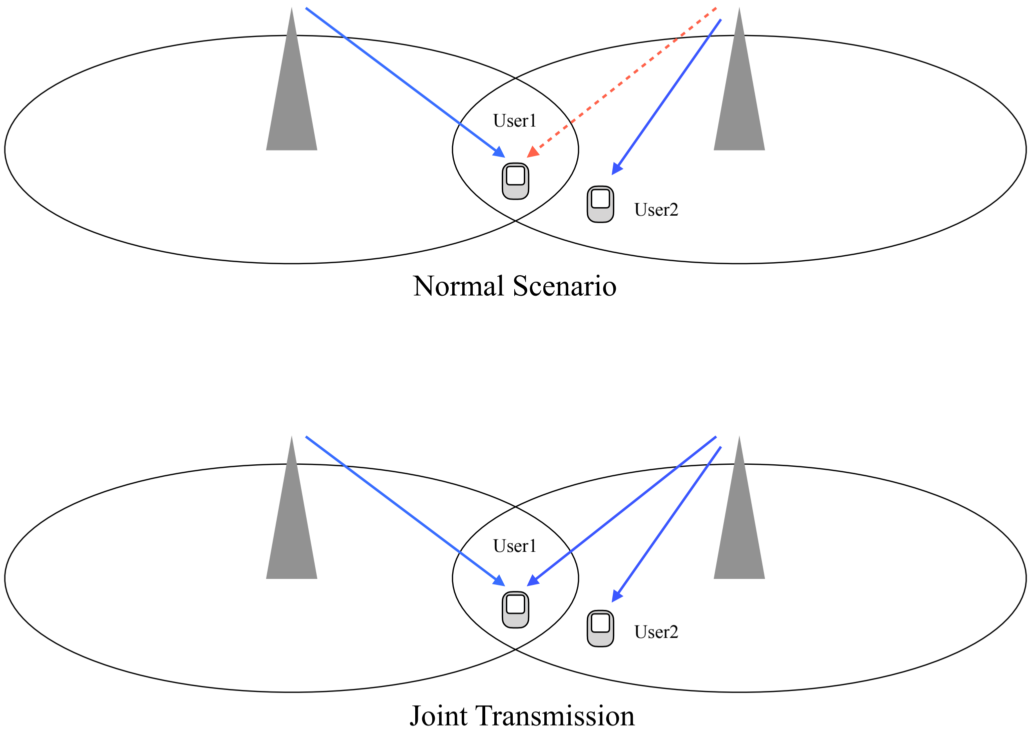


Fig. 2.3 Joint transmission.

Basic Concepts

- CoMP Cooperating Set

A set of different transmission location points transmit data to a single user in the time-frequency resource [64]. It should be noted that these points might be directly or indirectly involved in data transmission. Direct participation means the transmission points will transmit the signal to the subscriber in downlink; indirect participation means data will not go directly to the subscriber, but the location points will assist with scheduling/beamforming decisions.

• CoMP Transmission Points

Single or multiple transmission points transmit data to one user, and these transmission points are a subset of the CoMP cooperating set [64]. In CS/CB, one transmission point will transmit the signal to the subscriber; however, there is more than one transmission points from which to transmit the signal using JT. Because the transmission points are a subset of the cooperating set, they can be changed depending on the status.

• CoMP Measurement Set

A set of transmission points is measured and reported to the network system [64]. This set contains the channel state information related to the transmission points and the user. This measurement can support the system in making decisions (transmission points, scheduling, etc).

2.4.2 Literature Review in CoMP

In recent years, many studies have examined CoMP. Some of the challenges of CoMP include the triggering criteria, Signalling, the impact from cooperative BSs, and BS selection. Related work and discussion on this subject are shown after the challenge description.

Triggering Criterion of CoMP

The triggering criteria of CoMP are used to identify a scenario where performance does not match subscriber demand, and CoMP should be applied. However, when a large number of BSs and subscribers exist within a CN, CoMP cannot be used for all subscribers owing to high computational complexity. In JT, more than one BS transmits a signal to a subscriber, meaning BSs will face resource shortages. Therefore, the CoMP triggering criterion and ending time must be carefully considered; the system needs to make decisions based on limited or inaccurate reports from subscribers because of the dynamic environmental conditions. Furthermore, the advanced criterion requires complicated calculations, which might cause delays in the system. Since all triggering criteria have strengths and weakness, they could be deployed in different scenarios. Many studies have been conducted in this area.

• Literature Review in Triggering Criterion of CoMP

These triggering criteria differ in many articles. Some articles suggest that CoMP is triggered by subscriber's SINR or reference received signal power, while others suggest that it can be triggered by estimating subscriber data rates or total cell throughput.

The simplest way to determine CoMP processes is to measure current subscriber SINR. In [72], the JT are triggered by two criteria: the channel quality between the current serving BS and subscriber below the preset threshold or a high-speed data rate downlink requested by the subscriber. The equation of trigger criteria is proposed in [73], which is $10log(P_i/P) \leq \alpha$, where α is the preset value.The P_i represents the interference from cooperative BS and P represents the received power from the serving BS. If the above equation can be satisfied, it will execute JT to improve downlink performance. A similar method is proposed in [74] to estimate the ratio between received signal power and other interference.

In [75], it proposed that the JT can be triggered if the referenced received signal power between the serving BS and cooperative BSs is larger than a preset value. These criteria help the system to identify the best cooperative BS if the subscriber is located in a multi-overlapping area that it is served by more than one BSs. Furthermore, this work also proposes that this type of subscriber is suffering severe inter-cell interference and JT has significantly contributes to improved cell-edge user performance.

In [76], it is suggested that such triggering criteria should bring throughput into consideration, as shown in Equation 2.8. The T_{s1_o} and T_{s2_o} represent the original throughput in BS 1 and BS 2, and T_{s1_c} represents the throughput of BS 1 after JT has been triggered.

$$\alpha(T_{s1_o} + T_{s2_o}) \leq T_{s1_c} \tag{2.8}$$

This method has strong flexibility, meaning network operators can adjust the α to balance the network and cell-edge throughputs. A large value of α lets the system obtain a high network throughput and vice-versa. Other studies consider the possibility of using cell-edge performance only. In [77], it is suggested that trigger criteria facilitate this consideration. When the actual downlink performance is below the average system downlink performance, JT will be executed to improve upon this result.

Load and Scheduling

The status of cooperative BSs could cause several impacts on CNs. In detail, irregular BS deployment and high-density PC deployment in real CN will cause uneven load distribution. As JT occurs within two or three BSs, cooperative BS selection is crucial. In general, received signal strength in cell-edge is especially poor. Thus, incorrect cooperative BSs will provide little improvement of cell-edge performance. Because the centre subscribers have superior spectrum efficiency compared with cell-edge subscribers, BS load status could produce throughput loss when JT is applied in high-load BSs [76]; the performance for cell-edge subscribers could be improved, but the loss of overall network throughput will outweigh that improvement. Therefore, it is crucial to choose a suitable BS to participate in JT.

• Literature Review in Load and Scheduling

Base station information is an essential component in the BS selection process in JT [78]. When under the pressure of high volume information exchanges among BSs, the downlink CoMP gain will decrease with the increasing load. During high load situations, the information exchange among BSs will be significantly increased. If the information exchange is delayed, the JT gain will decrease.

The traditional resource allocation scheme is to divide resources into two parts: resources for centre subscribers and resources for JT subscribers. In [79], a method is proposed to balance the resource allocation between centre subscribers and cell-edge subscribers. Furthermore, the paper suggests that the resources for JT can be proportional to the number of cell-edge subscribers. When comparing JT and non-JT performance, the BS will serve the subscribers normally if there is no significant improvement. This method addressed the problem of sacrificing centre subscriber resources for JT.

The performance of JT is also investigated in [80], which examines two scheduling schemes: priority for JT subscribers and priority for non-JT subscribers. This scheme will schedule corresponding subscribers first and then consider the remaining subscribers. Simulations showed that the system will have a significant reduction if the priority JT scheme is applied in high-load scenarios, owing to the subscribers in BS centre experiencing superior channel conditions than those in the cell edge. In addition, moving directions must be considered. Priority non-JT schemes should be applied to the subscribers who are moving toward the BS centre as the channel conditions improve, meaning JT is not necessary for these subscribers. Thus, JT will not waste resources. Further research in [81] has also shown that this system will choose a scheme depending on the mobility of subscribers. Simulation results confirmed that performance can be markedly improved.

2.5 Summary

This chapter consists of the following major overview sections: EE in the CN, ML, traffic prediction in CN, JT in CoMP, and load balance.

In the reviews of the EE of CN, definitions and BS deactivation strategies were introduced. Where the EE optimisation problem took subscribers into consideration, owing to the large number of BSs and subscribers within the CN. Some studies only considered the EE and BSs, but ignored the overall signal quality of the network and resources usage. Meanwhile, the distance between BS and subscriber shall be put into consideration as well because this will bring impact on the network performance. Further investigation should consider the SINR of distribution where subscribers can still receive the optimal signal even if some BSs have been deactivated for improving the EE.

In the reviews of ML, it presents the basic concepts and advantages of probabilistic classifiers. Furthermore, the NBC was introduced and related works with traffic prediction in CN was shown. Based on the Bayes theorem, the methods were used for other types of prediction; however, they achieved high prediction and classification accuracy, but limited try in CN traffic prediction. However, some traditional and recent researches show that they only consider the traffic distribution, but ignore the detailed information from subscribers. Finally, studies did not fully utilise the traffic prediction result. In summary, the research shall be continues in traffic prediction with subscriber's information, and the way of combining traffic prediction result and energy saving.

In the reviews of JT, the problem is summarised in detail, including trigger criterion, load, and scheduling. The trigger criterion can significantly balance the overall network performance and cell-edge subscribers performance. Due to the double resources occupation in serving BS and cooperative BS, heavy-load BS should be carefully considered when the system decides to choose it for JT. The impact from executing JT in heavy-load BS shall be further investigated, and resources shortage issue in cooperative BS is lack of investigation. The resources allocation issue shall be solved to help the cooperative BS can use JT during high traffic demand period.

Chapter 3

Energy Saving with Optimal Cells Association in Heterogeneous Network

3.1 Introduction

The fast-growing traffic demand in contemporary CN brings a new challenge to network operators. The traditional network architecture cannot provide sufficient capacity, and network operators need to deploy increasing number of BSs to match the growing traffic demand. Due to the limited space, it becomes a challenge for high-density BS deployment. To tackle this challenge, network operators tend to change the network architecture, and the HetNet becomes a trend in the future network architecture. A set of PCs are deployed under MC's coverage to provide extra access points [82]. These low-power BSs will not only solve the challenge of insufficient deployment space but also significantly reduce the path loss. However, high-density PC deployment will increase the energy consumption, and it takes 30% of the operators cost [83]. The traffic fluctuation causes the load of BS is light during the midnight to the early morning, the network operators can switch off some light-load PCs to reduce energy consumption and OPEX [84].

3.1.1 Energy Efficiency in Cellular Network

The energy consumption of CN becomes a main challenge to the network operators. The increasing number of PCs are deployed in the HetNet, it can be seen that these PCs will increase the total network power consumption, increasing the OPEX of operating a CN. According to [85], BS can consume almost 80% of CN energy, and energy is mainly used for signal transmission, BS cooling and backhaul network. During light-traffic period, network only needs a small amount of BSs to serve the subscriber. Therefore, it will become a potential research area for improving EE in the HetNet. Although some of the functionalities might not be available in PC, switching off some PCs can significantly reduce the energy consumption, and BS sleep mode and corresponding BS switch-off strategy is proposed to reduce the energy consumption [86]. The beginning studies in switch-off strategy during light-traffic period are proposed in [87] [88]. The power reduction in different traffic demand and deployment scenarios are shown in these two researches, and it points out the tradeoff between power consumption and throughput. The energy saving in sleep mode is investigated in [89], and the research shows that the EE can be improved at light and medium load. Two switch-off algorithms are proposed to dynamically reconfigure BSs' working mode for energy saving in HetNet [90]. The BSs will be switched off when all subscribers can be transferred to neighbour BS, but research might ignore the subscribers might connect to remote BS, and the received signal strength will be weaker than the original serving BS. Moreover, PCs might not assign sufficient resources to subscriber if the load is heavy, degrading the QoS. Therefore, the QoS should be guaranteed before the switch-off decision is made, and spectrum efficiency should also be considered. The research [32] utilise integer convex optimisation to determine on-off pattern, the interference from neighbour BSs will be considered when the system makes the switch-off decision. The recent study in [91] proposes a two-stage optimisation, required resources is estimated and recorded by the system, and the result will be utilised in the second stage to calculate the minimum number of active BS.

Optimisation Challenge

When system decide to switch off some light-load BSs to achieve optimal EE, cell association, QoS shall be considered. The switch-off strategy should carefully consider the consequence when light-load PCs are switched off because the location of the subscribers is randomly distributed, subscribers might connect to remote BS. Switch-off strategy should avoid the problem of high path loss and worse SE. Therefore, EE problem can be formulated as an optimisation problem that the system shall switch off some light-load PCs, but the system should choose an optimal PCs combination to serve the subscribers. However, the optimisation problem will become a NP-hard problem with a large number of PCs and subscribers, and the computation complexity is high. To tackle this problem, optimisation problem will be divided into some sub-optimisation problems to reduce computation complexity. First, the power of PC is small, the interference from other MC coverage's PCs can almost consider as zero. Second, sub-optimisation can reduce the computation complexity because operators deploy small number of PCs within one MC coverage [92]. This chapter will apply SA to determine the best PCs combination when the system make the decision to switch off some PCs. The SA algorithm will compare the EE of new PCs combination with the previous PCs combination, algorithm will accept the new PCs combination if EE is larger than before. However, SA algorithm will accept the worse solution with a certain probability to avoid the local maximum value [93] [94]. After a certain iterations, the algorithm will determine the best PCs combination.

This chapter includes the following sections. Section two will describe the system model. Section three will formulate the optimisation problem and sub-optimisation problem. The SA algorithm will be described in detail, application of SA in PCs selection. Section four will show the simulation result of one MC and set of PCs scenario, including result in iteration, the best PCs combination and the influence from increasing load of PCs. Section five will show the simulation result of multiple MCs and PCs scenario. The last section will give the conclusion.

.

3.2 System Model

3.2.1 Optimisation Process

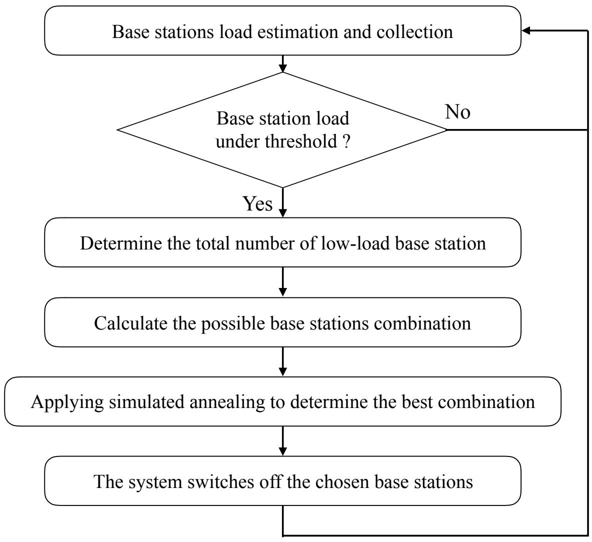


Fig. 3.1 Optimisation process.

The optimisation process of switch-off strategy is shown in Figure 3.1. All PCs shall upload their load information to the system no matter the load is higher or lower than the preset load threshold. At the beginning, the system will collect all BSs load information, and it will continues monitoring all PCs, which is above the preset load threshold. Otherwise, if the load of PCs is under the preset threshold, the system will count the total number of low-load PC. After the total number is determined, it will process to the next step to use SA algorithm to determine the best PCs combination, and which PCs shall be switched off. At the end of the algorithm, the best PCs combination will pass to the system, and the system will inform the corresponding PCs to prepare the switch-off strategy. After the PCs are switched off, the system will return to the first step to continuously monitor BS's load information.

3.2.2 Notation

Let's consider a HetNet with BSs deployment, multiple MCs and PCs. A set of PCs are deployed within the MC edge or high-traffic demand area to provide extra access points. Let's denote a set of MCs with M and PCs with P, where $m \in \{1,2,3,...,m\}$ and $p \in \{1,2,3,...,p\}$. If subscriber is connected to the MC or PC, the interference comes from the remaining BSs. Therefore, the interference BSs from MC tier and PC tier are denoted as $M^{'}$ and $P^{'}$. The subscribers i are randomly distributed within the network, and it is denoted as I, where $i \in \{1,2,3,...,i\}$.

3.2.3 Path Loss Model and SINR

It is assumed that the subscriber will connect to either MC or PC in the network. The subscriber will be served by only one BS, the serving BS is determined by the received signal strength, and the received signal strength of subscriber is estimated by path loss model.

Free Space Path Loss Model The free space path loss model is utilised in this chapter to estimate the loss in free space, the equation of free space path loss is shown in the following equation in terms of dB

$$
\begin{aligned}
FSPL &= 10log_{10}((\frac{4\pi df}{c})^2) \\
&= 20log_{10}(\frac{4\pi df}{c}) \\
&= 20log_{10}d + 20log_{10}f + 20log_{10}(\frac{4\pi}{c}) \\
&= 20log_{10}d + 20log_{10}f - 147.55
\end{aligned}
\tag{3.1}
$$

where d is measured in the units of metres and f is measured in the units of Hertz. c represents the speed of light. Therefore, according to the equation, if the transmission power of BS is known, the received signal strength of subscriber can be calculated.

SINR The *SINR* equation is shown by the following equation

$$SINR = \frac{S}{I+N} \tag{3.2}$$

where S represents the received signal strength at subscriber location, which is estimated by the above free space path loss model. I represents the interference from the remaining BSs. N represents the background noise.

Therefore, according to the Equation 3.2, the *SINR* of subscriber i, which is connected to MC m or PC p can be represented by

$$SINR_{im} = \frac{P_{im}G_{im}}{\sum_{m=1}^{M'} P_{im}G_{im} + \sum_{p=1}^{P'} P_{ip}G_{ip} + \sigma^2} \tag{3.3}$$

$$SINR_{ip} = \frac{P_{ip}G_{ip}}{\sum_{m=1}^{M'} P_{im}G_{im} + \sum_{p=1}^{P'} P_{ip}G_{ip} + \sigma^2} \tag{3.4}$$

where P_{im} and P_{ip} represent the transmission power of MC and PC. G_{im} and G_{ip} are the channel gain when the subscriber connects to the corresponding BS, which includes path loss only. σ^2 denotes the noise power. Therefore, the *SINR* of any subscriber anywhere within the network can be obtained, which will be used to estimate the downlink performance and to calculate the throughput.

3.2.4 Traffic and Load Model

In this chapter, it is assumed that the subscriber requires downlink service from the network. Let's denote the downlink service of subscriber i as D_i^τ, where τ is the traffic type indicator.

Three traffic types are categorised in this chapter, D_i^1, D_i^2 and D_i^3 . The specific downlink requirement in each type is 200Kbps, 400Kbps and 600Kbps.

The load model is based on the model presented in [89]. Let's denote the load of MC m as L_m, which is shown by the following equation

$$L_m = \frac{\sum_{i=1}^{I} R_{im} X_{im}}{R_{tm}} \tag{3.5}$$

$$where \quad R_{im} = \frac{D_i^\tau}{log_2(1 + SINR_{im})} \tag{3.6}$$

where R_{im} and R_{tm} represent the allocated resource of subscriber i and total resource number of MC. X_{im} will equal to 1 when subscriber i is connected to MC m, otherwise 0. The $SINR_{im}$ represent the $SINR$ of subscriber i when it is connect to MC, the calculation is shown in Equation 3.3. Therefore, the load of PC can be represented by the Equation 3.7.

$$L_p = \frac{\sum_{i=1}^{I} R_{ip} X_{ip}}{R_{tp}} \tag{3.7}$$

$$where \quad R_{ip} = \frac{D_i^\tau}{log_2(1 + SINR_{ip})} \tag{3.8}$$

Total resource number of PC p is denoted as R_{tp}. R_{ip} represents the allocated resources. X_{ip} will equal to 1 when subscriber i is connected to PC p, otherwise 0. $SINR_{ip}$ is shown in Equation 3.4.

3.2.5 Throughput of the System

According to the Shannon channel capacity formula [95] [96], the data rate of subscriber can be estimated by the following equation

$$C = B \times \log_2(1 + SINR) \tag{3.9}$$

where B represents the bandwidth. Let's define the throughput of subscriber as T_{im} and T_{ip} when subscriber connects to the corresponding BS tier. Therefore, subscriber's throughput can be calculated by the following equations

$$T_{im} = X_{im} R_{im} log_2(1 + SINR_{im}) \tag{3.10}$$

$$T_{ip} = X_{ip} R_{ip} log_2(1 + SINR_{ip}) \tag{3.11}$$

where X_{im} and X_{ip} will be equal to 1 when subscriber i is connected to the MC m and PC p, 0 otherwise. R_{im} and R_{ip} are the allocated resources of subscriber when subscriber is connected to the corresponding BS tier. The $SINR_{im}$ and $SINR_{ip}$ represent the $SINR$ when subscribers are connected to MC or PC respectively.

It is assumed that the throughput of the network is the sum of all subscriber's throughput within the CN. Therefore, according to the above throughput equations, the total throughput of the network can be obtained as follows:

$$T_{total} = T_M + T_P \tag{3.12}$$

$$\begin{aligned} where \quad & T_M = \sum_{m=1}^{M} \sum_{i=1}^{I} X_{im} R_{im} log_2(1 + SINR_{im}), \\ and \qquad & T_P = \sum_{p=1}^{P} \sum_{i=1}^{I} X_{ip} R_{ip} log_2(1 + SINR_{ip}) \end{aligned}$$

where T_M and T_P represent the total throughput in MC tier and PC tier.

Let's put Equations 3.6, $D_i^{\tau} = R_{im} * log_2(1 + SINR_{im})$ and Equations 3.8, $D_i^{\tau} = R_{ip} * log_2(1 + SINR_{ip})$ into the above Equations $T_M = \sum_{m=1}^{M} \sum_{i=1}^{I} X_{im} R_{im} log_2(1 + SINR_{im})$ and $T_P = \sum_{p=1}^{P} \sum_{i=1}^{I} X_{ip} R_{ip} log_2(1 + SINR_{ip})$. Therefore, the above equation of total throughput of CN can be expressed by the following equation

$$T_{total} = \sum_{m=1}^{M}\sum_{i=1}^{I} X_{im} D_i^{\tau} + \sum_{p=1}^{P}\sum_{i=1}^{I} X_{ip} D_i^{\tau} \ = D_i^{\tau} (\sum_{m=1}^{M}\sum_{i=1}^{I} X_{im} + \sum_{p=1}^{P}\sum_{i=1}^{I} X_{ip}) \tag{3.13}$$

It can be seen that the total throughput of network mainly depends on the total number of subscriber that CN can serve. Therefore, if the network has an optimal spectrum efficiency, the subscriber's service demand can be satisfied even the BS assigns small amount of resources.

3.2.6 Power Consumption Model

The power consumption of network can be divided into many parts. However, this chapter will only considers the power consumption of BS, and the power consumption for signal transmission will be the main focus. The power consumption model will consider two modes, which are active mode and sleep mode, and it is assumed that the BS will only stay in either of one mode in the simulation.

Active Mode Let's consider the total power consumption of BS in active mode, and same type BS is operated in the same power. The power consumption model is followed the model in [95]. It is assumed that the total network power consumption equals to the sum of BS transmission power. When all MCs and PCs stay in active mode, the total power consumption of the network can be represented by the following equation

$$P_{total} = \sum_{m=1}^{M} P_{tm} + \sum_{p=1}^{P} P_{tp} \tag{3.14}$$

where P_{tm} and P_{tp} represent the transmission power of MC and PC.

Sleep Mode The deeper of the sleep mode is executed, the less BS power consumption [95]. Although BS still consumes a small amount of power in sleep mode, this power consumption can be negligible and considered as zero. In this chapter, PCs receive switch-off message from the system, these PCs will turn into switch-off mode, and the power consumption is considered as zero. Moreover, switch-off strategy will be only applied in PC tier, the MC will be operated in regular power to provide coverage.

3.2.7 Energy Efficiency Model

Let's defines the E as the EE of network. EE is defined as the total network throughput divided by the total energy consumption of network, which is calculated as bits per joule [84] [95] [97]. Let's combine Equation 3.13 and Equation 3.14, the EE of network can be represented by the following equation

$$E = \frac{T_{total}}{P_{total}} = \frac{\sum_{m=1}^{M}\sum_{i=1}^{I} X_{im} D_i^{\tau} + \sum_{p=1}^{P}\sum_{i=1}^{I} X_{ip} D_i^{\tau}}{\sum_{m=1}^{M} P_{tm} + \sum_{p=1}^{P} P_{tp}} \tag{3.15}$$

Therefore, the objective is to maximise the EE of CN. According to the above equation, it can be seen that the EE can be improved by either improving the total network throughput or reducing the total network energy consumption.

3.3 Optimisation of Energy Efficiency

3.3.1 Problem Formulation

According to the Equation 3.15, the objective of the optimisation is to maximise the value of equation, which is shown below

$$max(E) = max(\frac{\sum_{m=1}^{M}\sum_{i=1}^{I} X_{im} D_i^{\tau} + \sum_{p=1}^{P}\sum_{i=1}^{I} X_{ip} D_i^{\tau}}{\sum_{m=1}^{M} P_{tm} + \sum_{p=1}^{P} P_{tp}}) \tag{3.16}$$

According to the system process, the system will switch off some low-load PCs to reduce the energy consumption, and it shall determine which PC should stay in active mode from the existing PCs. When some PCs are switched off, the subscriber's $SINR$ will be changed because the serving BS might not be the same with the original one. It is assumed that n PCs are under the load threshold in the network, and the system will switch off same number of PCs to reduce energy consumption. Therefore, the system needs to select n PCs out of P, where $n < P$. It can be seen that the total number of possible PCs combination is C_P^{P-n}, and the system needs to estimate EE with each possible PCs combination and determine the best PCs combination. The problem can be considered as a combination problem that system needs to choose best PCs combination to maximise EE of CN with limited resources.

The total throughput $T_{total}^{'}$ and total power consumption $P_{total}^{'}$ equations are different when n PCs are switched off, the equations can be represented as

$$T_{total}^{'} = \sum_{m=1}^{M}\sum_{i=1}^{I} X_{im}D_i^{\tau} + \sum_{p=1}^{P-n}\sum_{i=1}^{I} X_{ip}D_i^{\tau} \tag{3.17}$$

$$P_{total}^{'} = \sum_{m=1}^{M} P_{tm} + \sum_{p=1}^{P-n} P_{tp} \tag{3.18}$$

Let's combine the Equation 3.17 and 3.18, the new EE equation can be obtained, and it will be the cost function in the optimisation problem, which is represented as

$$\begin{aligned} max \quad f(x) &= \frac{\sum_{m=1}^{M}\sum_{i=1}^{I} X_{im}D_i^{\tau} + \sum_{p=1}^{P-n}\sum_{i=1}^{I} X_{ip}D_i^{\tau}}{\sum_{m=1}^{M} P_{tm} + \sum_{p=1}^{P-n} P_{tp}} \\ where \quad X_{im}D_i^{\tau} &= X_{im}R_{im}log_2(1+SINR_{im}) \\ X_{ip}D_i^{\tau} &= X_{ip}R_{ip}log_2(1+SINR_{ip}) \end{aligned} \tag{3.19}$$

$$st. \begin{cases} X_{im} + X_{ip} = 1, \forall_m \in M, \forall_p \in P, \forall_i \in I \\ \sum_{i=1}^{I} R_{im}X_{im} \leqslant R_{tm}, \forall_m \in M, \forall_i \in I \\ \sum_{i=1}^{I} R_{ip}X_{ip} \leqslant R_{tp}, \forall_p \in P, \forall_i \in I \\ \eta R_{tp} \leqslant \sum_{i=1}^{I} R_{ip}X_{ip}, \forall_p \in P, \forall_i \in I \end{cases} \tag{3.20}$$

where X_{im} and X_{ip} equal to 1 if subscriber is connected to the corresponding MC and PC, otherwise 0. The equation of $X_{im} + X_{ip} = 1$ represents subscriber will connect to either MC or PC. The equation $\sum_{i=1}^{I} R_{im}X_{im} \leqslant R_{tm}$ and $\sum_{i=1}^{I} R_{ip}X_{ip} \leqslant R_{tp}$ represent that total allocated resources can not larger than the total resources of MC and PC, and the value of R_{tm} and R_{tp} are depend on the system bandwidth. $\eta R_{tp} \leq \sum_{i=1}^{I} R_{ip}X_{ip}$ represents that PC load should be larger than threshold η, otherwise PC will be switched off.

Sub-Optimisation Problem The optimisation problem become a NP-hard problem because the total number of PCs and subscriber are enormous in the real network, but the optimisation problem will not be difficult to determine the best PCs combination in small scale. Due to the difficulty of solving the optimisation problem, optimisation problem can be decomposed into some sub-optimisation problems. According to the scenario, sub-optimisation problem will be considered as the scenario that a set of PCs are deployed within one MC coverage. Due to the PC's transmission power will not be set in high level, the transmission distance is limited. Therefore, one group of PCs will not cause significant interference to the other PCs group. Meanwhile, major interference will come from MCs, but the parameters of MCs will not be changed in most of the situation. Therefore, the interference from MCs can be considered as a constant. Therefore, the PCs can be divided into many group to reduce the total number of solution. It can be seen that if there are p PCs deployed within the CN. The total number of solutions will become C_p^{p-n} if n PCs shall be switched off. It is assumed that the PCs can be evenly divided into m group the total number of PCs in each group is p/m, the total solutions number will become $C_{p/m}^{(p/m)-n}$. Compare C_p^{p-n} with $C_{p/m}^{(p/m)-n}$, it can be seen that the total number of solution can be significantly reduced. The time for finding the optimal solution can be reduced.

To achieve the global solution, the system needs to determine the best PCs combination in each sub-optimisation first, and the best network's PCs combination will equal to the sum of the PCs combination in each PCs group. Therefore, m groups of PCs can be obtained, where $\{P_1, P_2, P_3 ... P_m\} \in P$, $\sum_{m=1}^{M} P_m = P$. The subscribers belong to the corresponding MC and PCs group are defined as $\{I_1, I_2, I_3 ... I_m\} \in I$ and $\{I_{p1}, I_{p2}, I_{p3} ... I_{pm}\} \in I$, where I_m represents the subscriber is served by MC m. I_{pm} represents the subscriber is under the coverage of

MC m and belongs to the PCs group p, and $\sum_{m=1}^{M} I_m + \sum_{m=1}^{M} I_{pm} = I$. It is assumed that the switch-off PCs number is n in one PCs group, the sub-optimisation problem can be represented as

$$\begin{aligned} max \quad f(x) &= \frac{\sum_{i=1}^{I_m} X_{im} D_i^{\tau} + \sum_{p=1}^{P_m - n} \sum_{i=1}^{I_{pm}} X_{ip} D_i^{\tau}}{P_{tm} + \sum_{p=1}^{P_m - n} P_{tp}} \\ where \quad X_{im} D_i^{\tau} &= X_{im} R_{im} log_2(1 + SINR_{im}) \\ X_{ip} D_i^{\tau} &= X_{ip} R_{ip} log_2(1 + SINR_{ip}) \end{aligned} \tag{3.21}$$

$$st. \begin{cases} X_{im} + X_{ip} = 1, \forall_m \in M, \forall_p \in P, \forall_i \in I \\ \sum_{i=1}^{I_m} R_{im} X_{im} \leqslant R_{tm}, \forall_m \in M, \forall_i \in I \\ \sum_{i=1}^{I_{pm}} R_{ip} X_{ip} \leqslant R_{tp}, \forall_p \in P, \forall_i \in I \\ \eta R_{tp} \leqslant \sum_{i=1}^{I_{pm}} R_{ip} X_{ip}, \forall_m \in M, \forall_p \in P, \forall_i \in I \end{cases} \tag{3.22}$$

3.3.2 NP-Hard Proof

To prove the above optimisation problem is NP-hard, it shall prove that a relaxed instance of the above optimisation is similar to the Capacitated Facility Location Problem (CFLP). The original problem is summarised as follows. A set of the sites and a set of the customers that denoted as $s \in \{1, 2, 3, ..., s\}$ and $c \in \{1, 2, 3, ..., c\}$, and the customers shall be served by the sites. Moreover, it denotes the cost of opening a site as Co_s and cost $Co_{c,s}$ of assigning a customer c to a site s. The capacity of each site is limited, and the maximum value is denoted as Max_s. The CFLP problem is to use minimum cost to satisfy all customers demands under the site's capacity, and the CFLP problem is proved as a HP-hard problem [98] [99] [100].

Considering the optimisation in this chapter, it assumed that all subscribers should be served, and the relaxed optimisation problem reduces to finding one set of BSs that can serve all subscribers when total resources of one BS Max_s is limited. The cost of operating one BS is Co_s, and the resource for serving subscriber c by BS s is $Co_{c,s}$. This problem is the same with NP-hard problem, CFLP. Therefore, the above optimisation problem is NP-hard.

3.3.3 Simulated Annealing

The SA algorithm is inspired by the annealing process in metallurgy. Annealing is a heating and cooling process to get a certain size material and reduce product defects, the process will gradually reduce temperature to control thermal equilibrium in each stage in order to obtain the best structure [101]. SA algorithm is shown in the Algorithm 1. The benefit of using SA is due to the flexibility that SA can do with arbitrary systems and cost functions. Meanwhile, its ability to find the global optimal solution will be the second reason of using SA to obtain the best PCs combination, this algorithm will not stuck in any solution, the advantage of this algorithm is the ability of jumping out from the local optimal solution. Moreover, the algorithm has the strength of versatile because SA is not rely on any restrictive properties of any model.

Algorithm 1: Simulated Annealing

- Initialise temperature T, and set $S_{best} = S_{random}$ and $R_{best} = 0$

for $i = 1$ **do**
  Choose random solution $S_{current}$ and calculate result $R_{current}$
  Calculate δ = $R_{current}$ - R_{best}
  if δ < *0* **then**
    $S_{best} \leftarrow S_{current}, R_{best} \leftarrow R_{current}$
    else
      Generate a Q, a random number between 0 and 1
      if Q < *exp (-δ / T)* **then**
        $S_{best} \leftarrow S_{current}, R_{best} \leftarrow R_{current}$
        else
          Reject $S_{current}$ and $R_{current}$
  Decrease the temperature $T = T * Ratio_{decrease}$
  $i = i + 1$
end
Algorithm will be stopped when temperature reach minimum value or meet some stopping criteria

At the beginning of algorithm, it will set a temperature T and generate a random solution S_{random} for cost function, and the algorithm let the best solution S_{best} equal to S_{random}, the corresponding value of cost function is marked as R_{best}. At the first iteration, it will set the iteration number as $i = 1$ and begin the SA algorithm. The second solution will partially change the previous solution, and the second solution will be marked as $S_{current}$, and the corresponding cost function value will be calculated, marked as $R_{current}$. Algorithm will accept and record the solution if the value of cost function is smaller than the previous solution's value. If the value is larger than the previous solution's value, algorithm will accept the solution if the criterion $exp(-\delta/T) > Q$ is satisfied, where δ is the difference between $S_{current}$ and S_{best}. Q is a random variable between (0, 1). If the criterion is satisfied, it will marked S_{best} = $S_{current}$ and R_{best} = $R_{current}$. The reason of the algorithm will accept a worse solution under a certain probability is to avoid the local maximum or minimum value that it might stop the algorithm to find the global maximum/minimum value. According to the expression $exp(-\delta/T)$, the larger value of T, the algorithm will have a higher chance to accept worst solution. Because the T will be decreased, the accepting chance will become smaller. At the end of the iteration, the temperature will be reduced according to the $Ratio_{decrease}$ and repeat the above algorithm. The process will repeat until the temperature reaches minimum value or some stopping criteria are satisfied.

3.3.4 Cell Selection via Simulated Annealing

Solutions generation

Different PCs combination will become the solutions in SA, and the performance of each PCs combination will be tested in each iteration. In two consecutive iteration, only part of the PCs status will be changed, closed to opened or opened to closed. The following part will show how the PC status is changed between two PCs combinations.

One Pico cell switched off In the case of only one PC shall be changed to close status in one iteration. All PCs combinations are generated as follows. After the first PCs combination is chosen, for example, $1,2,3,4,...,p-1$, which is shown in the figure a) of Figure 3.2, and the red colour PC p means the corresponding number PC will not be selected, and it will be set as closed status. In the next iteration, the next PCs combination will be changed based

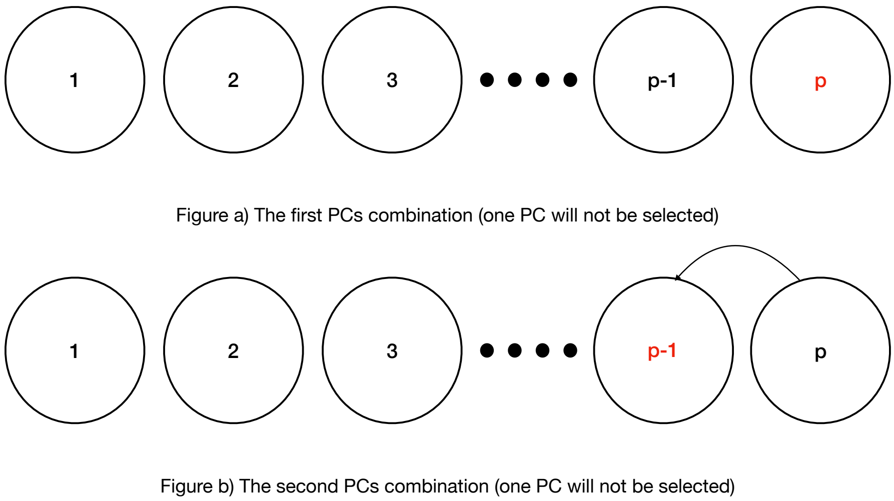


Fig. 3.2 Solutions generation. (one PC will not be selected)

on the first PCs combination, the last PC p will be chosen and changed the status to opened, and the second of the last PC $p-1$ will not be selected and changed the status to closed. Therefore, the second PCs combination will become $1,2,3,4,...,p-2,p$, which is shown in the figure b) of Figure 3.2. As you can see the number will jump to the previous PC in each iteration, and this process will be continues until the number is decreased to the first number, which is 1, and the PCs combination is $2,3,4,...,p$. By applying the above method, all the possible PCs combinations can be tested in each iteration.

Multiple Pico cells switched off In the case of more than one PC will not be selected in each PCs combination, the PCs combinations will be tested group by group. The solutions generation in each group is shown as follows. For example, if three PCs will not be selected in each SA iteration, the first group of PCs combinations are shown in Figure 3.3 from figure a) to d). The first PCs combination is chosen as what it shows in figures a) of Figure 3.3, and the last three PCs will not be selected, which is marked as red colour. After the first PCs combination is tested in SA. The second PCs combination will be selected as what it shown

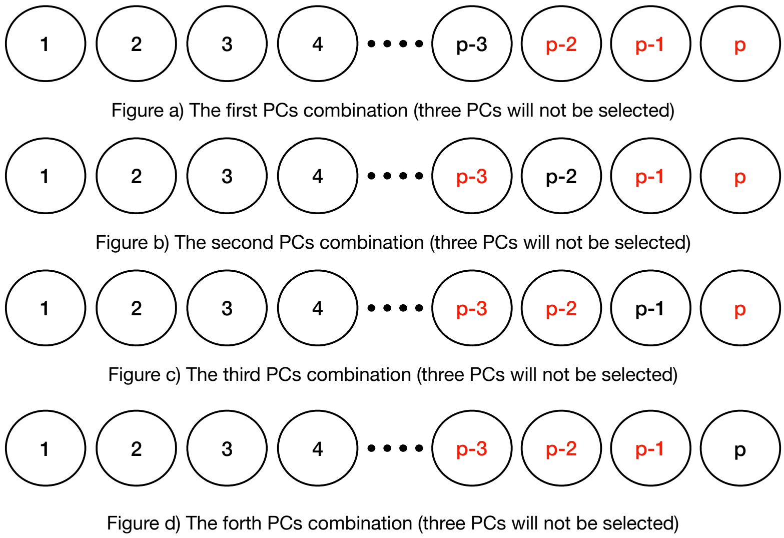


Fig. 3.3 Solutions generation. (more than 1 PC will not be selected)

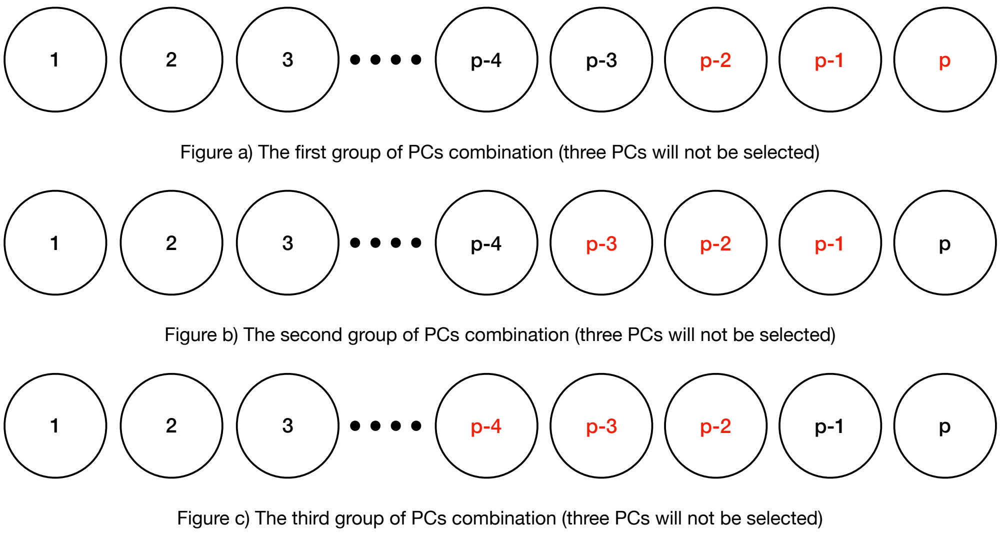


Fig. 3.4 Solutions switching.

in figure b) of Figure 3.3. The figure b) shows that the status of PCs $p-3$ and $p-2$ are both changed in this PCs combination, the PC $p-3$ is changed from opened to closed, and PC

$p-2$ is changed from closed to opened. The remaining PCs are maintain the same status with the previous PCs combination. The PCs combinations in figures c) and d) show how the next two PCs combinations will be changed, both PCs combinations only change two PCs status in each PCs combination. After the SA test the performance of PCs combination in figure d), the solution will continue to test the following group of PCs combinations. The process of changing from one group of PCs combination to another group is shown in Figure 3.4, the figure a) shows first PCs combination in the previous group, which is the same as figure a) of Figure 3.3. The figure b) of Figure 3.4 shows the first PCs combination of next group, as what it shows in figure b), the starting PC is $p-3$. The all three PCs will more forward again after the current group of PCs combinations are all tested. As what the figures shown in Figure 3.4, three red colour PCs are all moved one number forward. After the second group of PCs combinations are tested, it will repeat the same process, and all three PCs will move again in the next group, which is shown in figure c) of Figure 3.4. This process will continue until the starting PC number reach 1, and all the possible PCs combinations can be tested in SA.

Simulated annealing for cell selection algorithm

Applying the SA algorithm in PCs selection to determine the best PCs combination, the modified algorithm is shown in Algorithm 2. Let's set the total number of PC as p, and it is assumed that there are n PCs in low-load status after the system collects load information from the network, satisfying the condition that $L_p < \eta$. Therefore, the total possible number of PCs combinations is C_p^{p-n}, and each PCs combination can be obtained by applying the above PCs combination generation method.

The system chooses the first solution from C_p^{p-n} and estimates the EE of the PCs combination, denoted as $E_{current}$. The algorithm will set the initial temperature, which is denoted as T.

Algorithm 2: Cell Selection via Simulated Annealing

• Set n = 0
if $L_p < \eta$ **then**
    $n = n+1$
• Calculate the total number of n
• Choose a solution from C_p^{p-n}, marked as $S_{current}$, and calculate the corresponding energy efficiency, marked as $E_{current}$
• Set the initial temperature T
• Set $E_{best} = E_{current}$, $S_{best} = S_{current}$
for $i = 1$ **do**
    Change the solution based on the solution algorithm and denote as $S_{current}$, set the corresponding EE as $E_{current}$
    if $E_{best} < E_{current}$ **then**
        $S_{best} \leftarrow S_{current}, E_{best} \leftarrow E_{current}$
    else
        Compute $\delta = E_{best} - E_{current}$, and $P = \exp(-\delta / T)$
    if $Q < P$, **then**
        $S_{best} \leftarrow S_{current}, E_{best} \leftarrow E_{current}$
    else
        Reject $S_{current}$ and $E_{current}$
    • $T = T * Ratio$
    • $i = i + 1$
end
Algorithm will be stopped when temperature reach minimum value or all Pico Cells combinations are tested

---

Due to the first test in the algorithm, it will set $S_{best} = S_{current}$ and $E_{best} = E_{current}$. After the first iteration is finished, the system will estimate EE with the remaining PCs combinations. By applying the above PCs combination generation method, the method will partially change the previous PCs combination, marked as $S_{current}$, and the algorithm will calculate the corresponding cost function value, marked as $E_{current}$. If the new $S_{current}$ can achieve a better EE value, algorithm will accept the current PCs combination and the corresponding EE value, denoted as S_{best} and E_{best}. However, if the value of $E_{current}$ is not higher than the recorded value E_{best}, the system will calculate the value of $\delta = E_{best} - E_{current}$ for the equation $P = exp(-\delta/T)$. The system will only accept $S_{current}$ and $E_{current}$ if and only if the value of P larger than a random variable Q, which is distributed between $(0,1)$. If the above condition is satisfied, it will set $S_{best} = S_{current}$ and $E_{best} = E_{current}$. After one iteration is finished, the algorithm will decrease the temperature T according to the preset decreased ratio in the algorithm. The algorithm will be stopped when all PCs combinations are tested or the temperature reach the minimum value.

3.4 Simulation Result with Single MC and Multiple PCs

3.4.1 Scenario

It is assumed that the scenario is set in a light-traffic demand period with a single MC deployment, and a set of PCs are deployed in MC edge to provide access points, which is shown in Figure 3.5. The x axis and y axis represent the size of the scenario, one pixel within the figure represents 800x800 square meter. The z axis represents the received signal strength from BSs. The colour is changed from yellow to deep blue, and the lighter colour represent better received signal strength. The peak within the figure will be location of the MC and SCs. In this scenario, 25% subscribers are uniformly distributed within MC coverage, and the remaining subscribers are uniformly distributed within the set of PCs coverage. Each traffic type takes 33% in the subscriber group, and subscriber's location is fixed during the

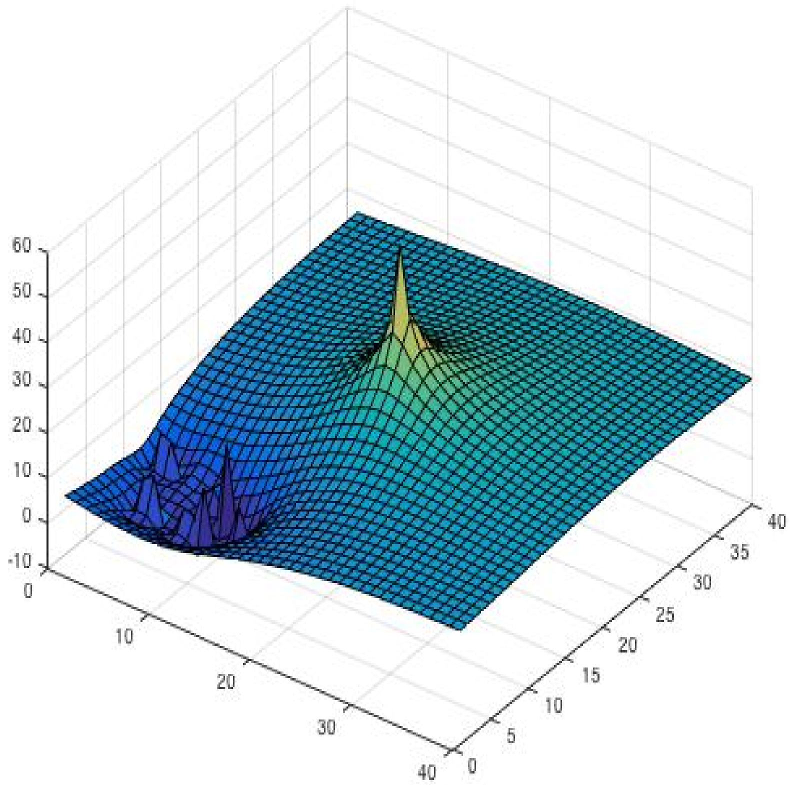


Fig. 3.5 Single macro cell and multiple pico cells scenario.

Table 3.1 Simulation parameters

Parameters	Value
Carrier Frequency	1800MHz
System Bandwidth	20MHz
Number of Resource Block	100
Load Threshold for Simulated Annealing	30%
Macro / Pico Cells Number	1 / 8
Macro Cell Transmission Power	46dBm
Pico Cell Transmission Power	24dBm
Noise Power	-174dBm/Hz
Path Loss Model	Free Path Loss Model
Traffic Type	200Kbps/400Kbps/600Kbps

simulation. The switch-off load threshold is set 30% to trigger the SA algorithm. The preset load threshold is set as 30% to trigger the SA algorithm to select the best PCs combination, and the other simulation parameters are shown in Table 3.1.

3.4.2 Simulation Result

Figure 3.6 shows the EE in each iteration when different PCs combination is chosen. At the beginning of SA, simulation result shows that some PCs combinations provide worse EE than the previous combination, but the algorithm still accept these PCs combinations when T is still a large value. However, the T is decreased with the increasing number of iteration, algorithm will have a smaller chance to accept worst PCs combinations. It can be seen that high value of T will let algorithm move out the local maximum area, which is shown in Figure 3.6 with the first and second peak. In each PCs combination, the subscriber's *SINR* might not be identical, and the allocated resources will be influence by the *SINR*, more subscribers can connect to the network if network has higher spectrum efficiency.

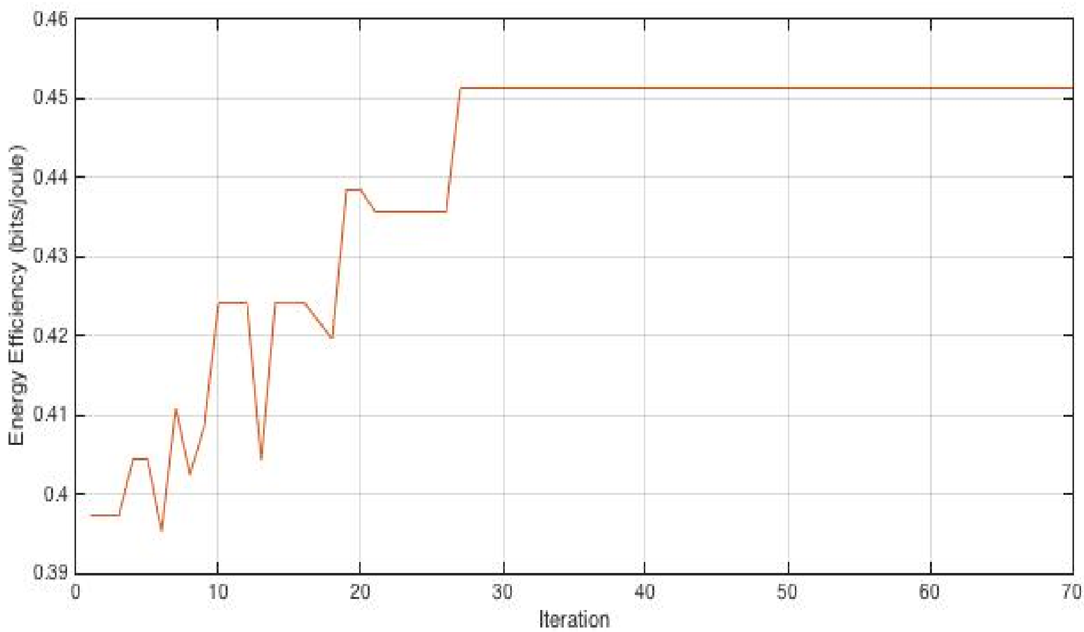


Fig. 3.6 Simulated annealing.

3.4.3 Pico Cells Combination

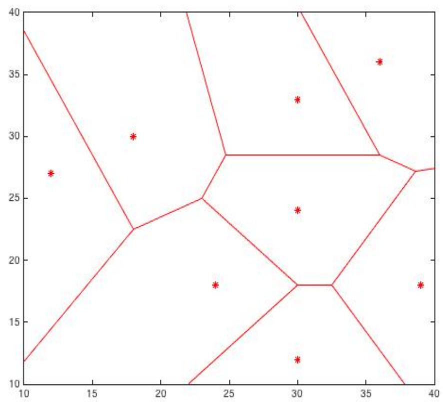


Fig. 3.7 Original active PCs.

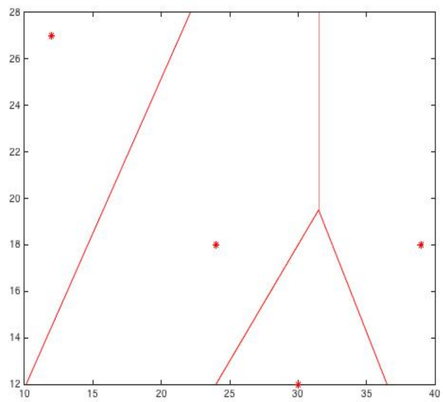


Fig. 3.8 Best PCs combination. (PCs : 2, 3, 6, 7)

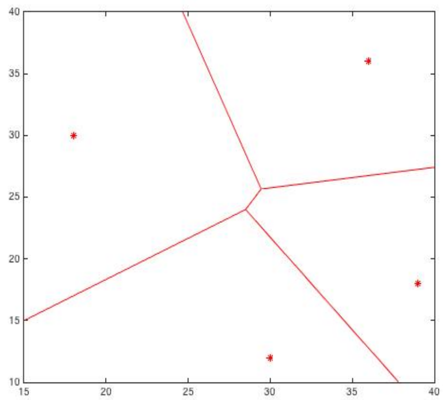


Fig. 3.9 PCs combination. (PCs: 1, 6, 7, 8)

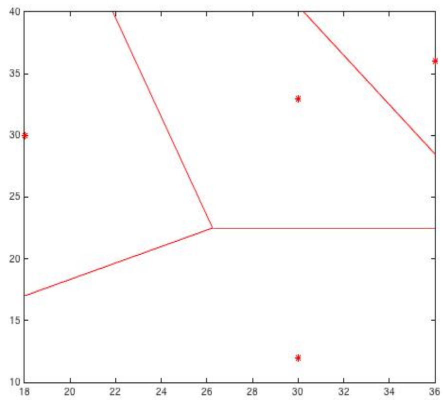
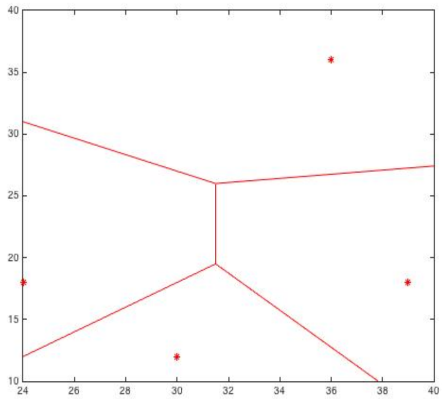


Fig. 3.10 PCs combination. (PCs: 1, 5, 6, 8)

After total number of switch-off PCs is determined by the system, 4 PCs are under the load threshold in the simulation. Therefore, C_8^4 kinds of PCs combination can be obtained. Figure 3.7 shows the original PCs deployment. Each EE will be calculated in the algorithm, part of the PCs combinations are shown in Figures 3.8-3.12. Compared with the original active PCs and other PCs combination, the simulation result shows that the best PCs combination is PCs 2, 3, 6 and 7, which is shown in Figure 3.8. The remaining figures show the other PCs combinations but it cannot provide higher EE than the PCs combination in Figure 3.8.



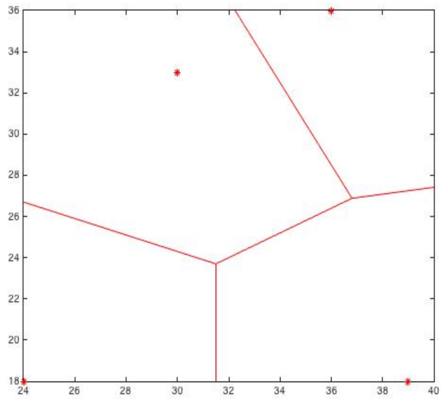


Fig. 3.11 PCs combination. (PCs: 1, 3, 6, 7) Fig. 3.12 PCs combination. (PCs: 1, 3, 5, 7)

3.4.4 Simulation with Increased Load

In this section, it will set up different traffic type users to simulate different load scenario. There are total three different type of user in the scenario, called type 1, type 2 and type 3. Larger number of the traffic type will require higher downlink speed within the simulation, and the traffic demand is defined as followed, type 1 subscribers will require 200kbps downlink during the simulation, and type 2 and type 3 subscribers will require 400kbps and 600kbps downlink speed. It sets total 120 subscribers within MC coverage, and it sets 33% of the subscribers in each traffic type. In this simulation, the number of type 3 subscriber will be slightly increased to represent the increasing traffic demand within the network, and it resets half of the type 1 subscribers to type 3. Therefore, the total traffic demand in downlink is increased. Figure 3.13 shows the EE in each iteration when SA is applied to determine the best PCs combination.

Based on the preset load threshold, the system will switch off 4 PCs. It can be seen that EE will be increased because of the higher resources utilisation under the condition of the same number of PCs are switched off, which is shown in Figure 3.13. The orange line has higher EE than the blue line at the end of the simulation, maximum value of EE will be obtained if all PCs reach 100% resources utilisation. The increased number of type 3 subscribers will force the system to utilise more resources to match the increasing traffic

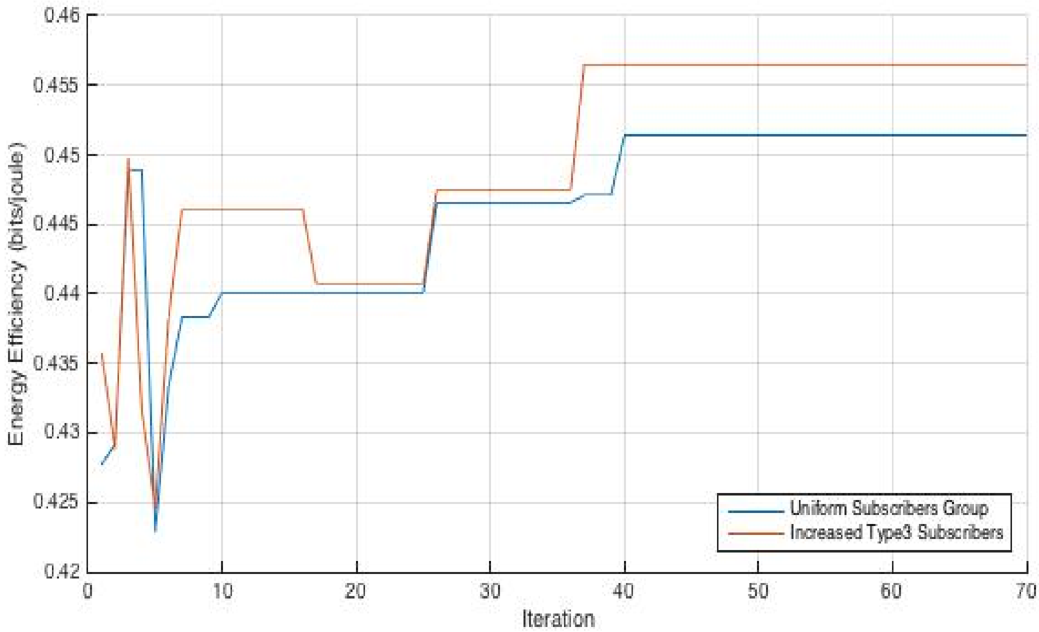


Fig. 3.13 Increase type 3 subscribers.

demand. After the selected PCs are switched off, the traffic is redistributed to the remaining PCs.

The comparison is made to show the resource utilisation, resource utilisation in three scenarios are shown in Figure 3.14, the original scenario with normal setting in BSs and subscribers, the scenario with uniform subscribers distribution in three traffic types and high percentage of traffic type 3 subscribers. Resources utilisation is shown in Figure 3.14 that the resources utilisation in original scenario is 44%. After 4 PCs are switched off, the resources utilisation is increased to 65% with uniform traffic type, the EE is increased due to the power for operating the BSs is reduced, but the total serving subscribers is not changed. In the last simulation, 69% resource utilisation with the increased number of type 3 subscribers. The higher resource utilisation can be obtained due to the empty resources within the network even some of the PCs are switched off, and these resources can be used for serving the increased traffic demand. The simulations in three different scenarios show that the PCs switch-off strategy can improve the EE of network and resources utilisation.

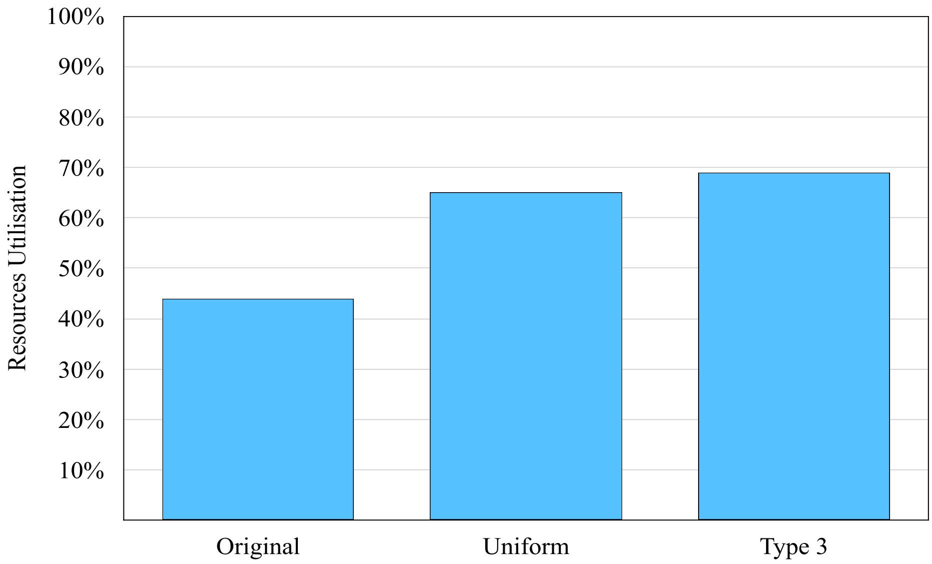


Fig. 3.14 Resources utilisation.

3.5 Simulation Result with Multiple MCs and PCs

3.5.1 Scenario

In this section, let's consider the scenario that four MCs and four sets of PCs are deployed within the scenario. Each PC set will contains 7 PCs, and one PC set will be deployed within one MC coverage, which is shown in Figure 3.15. In Figure 3.15, The x axis and y axis represent the size of the scenario, one pixel within the figure represents 1000x1000 square meter. The z axis represents the received signal strength from BSs. The colour is changed from yellow to deep blue, and the lighter colour represent better received signal strength. The peak within the figure will be location of the MC and SCs.the 33% subscribers are uniformly distributed within MC, the remaining 66% are uniformly distributed within the PC coverage. Subscribers require downlink service during the simulation, and the location is fixed during the simulation. The load threshold is set 30% to trigger the SA to switch off low-load PCs.

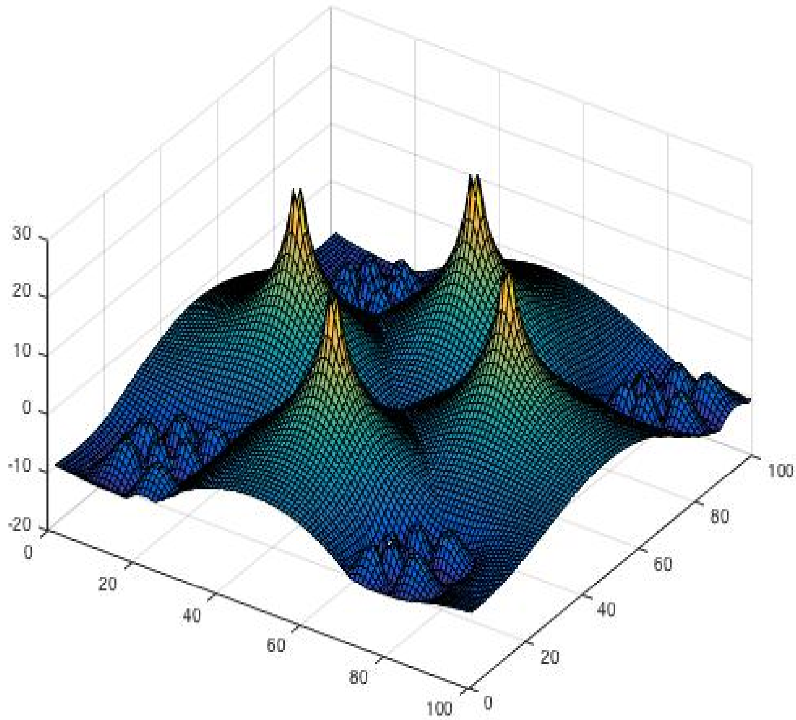


Fig. 3.15 Multiple macro cells and pico cells.

3.5.2 Simulation Result

Figure 3.16 shows the initial load status of PC. According to the preset load threshold for triggering SA, some PCs are under 30% load, which is PCs 2, 5, 8, 11, 12, 13, 16, 17, 20, 21 and 25. Based on the preset PCs group, the switch-off PC number in each group is 2 PCs in first group, 3 PCs in second group, 3 PCs in third group and 1 in last group. The simulations result in each group is shown in the following Figures 3.17 - 3.20, the figures show that the EE in each can be slightly improved.

3.5.3 Pico Cells Combination

Due to each MC will contains 7 PCs, the PCs are distributed as followed. From MCs 1-4, each MC will contain the number of 1-7, 8-14, 15-21 and 22-28 PCs. According to the above

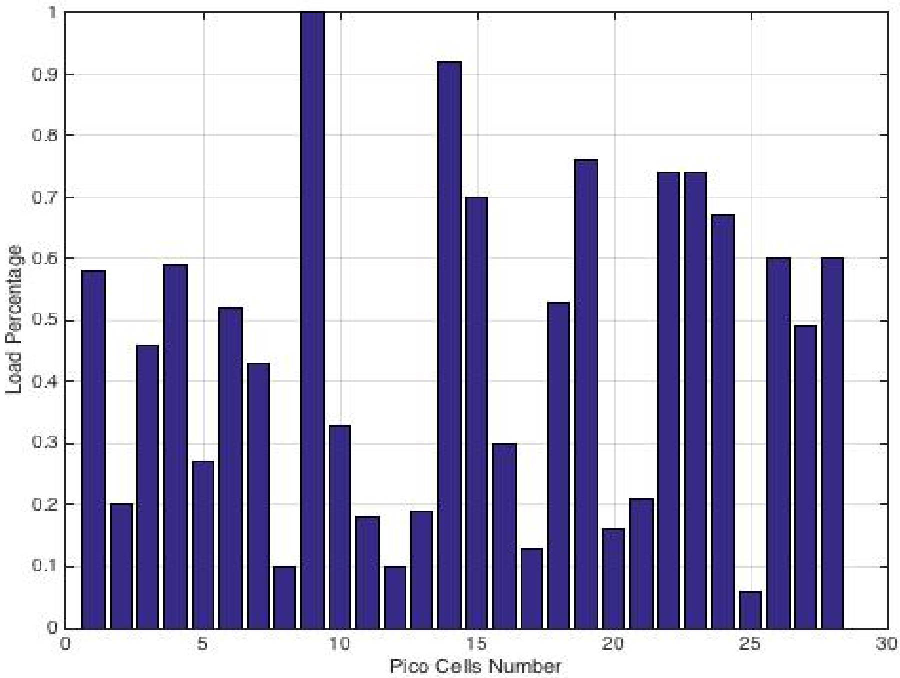


Fig. 3.16 Pico cells load status.

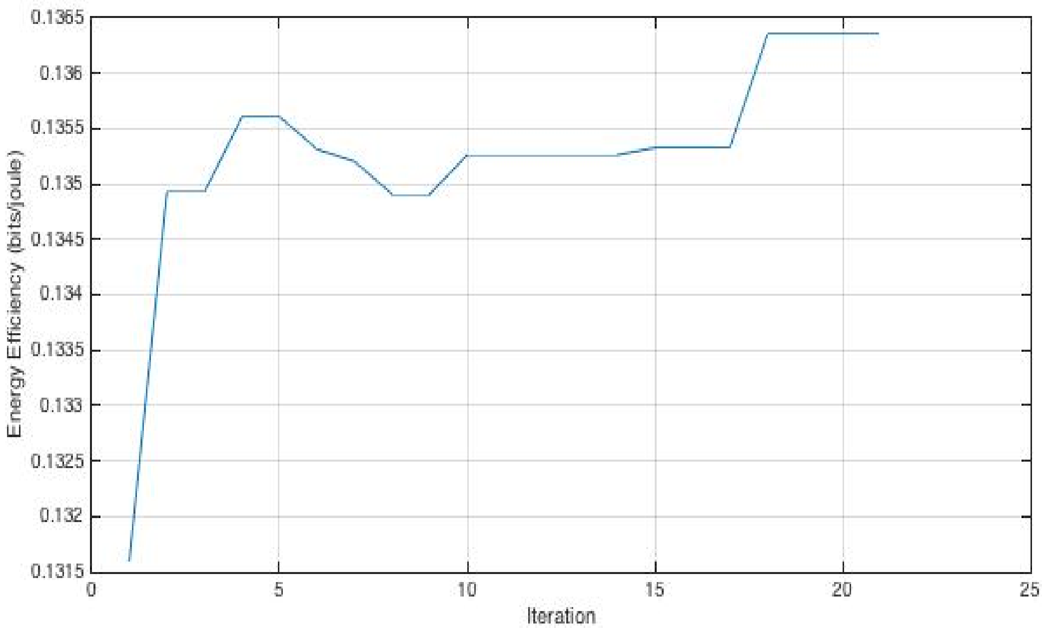


Fig. 3.17 Simulated annealing in macro cell 1 and pico cells 1-7.

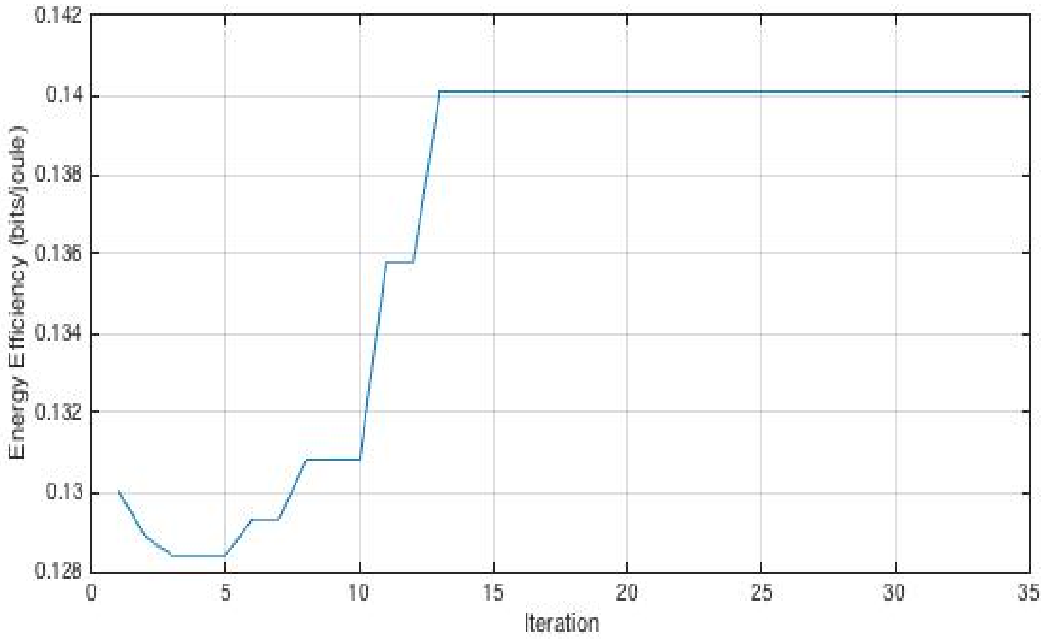


Fig. 3.18 Simulated annealing in macro cell 2 and pico cells 8-14.

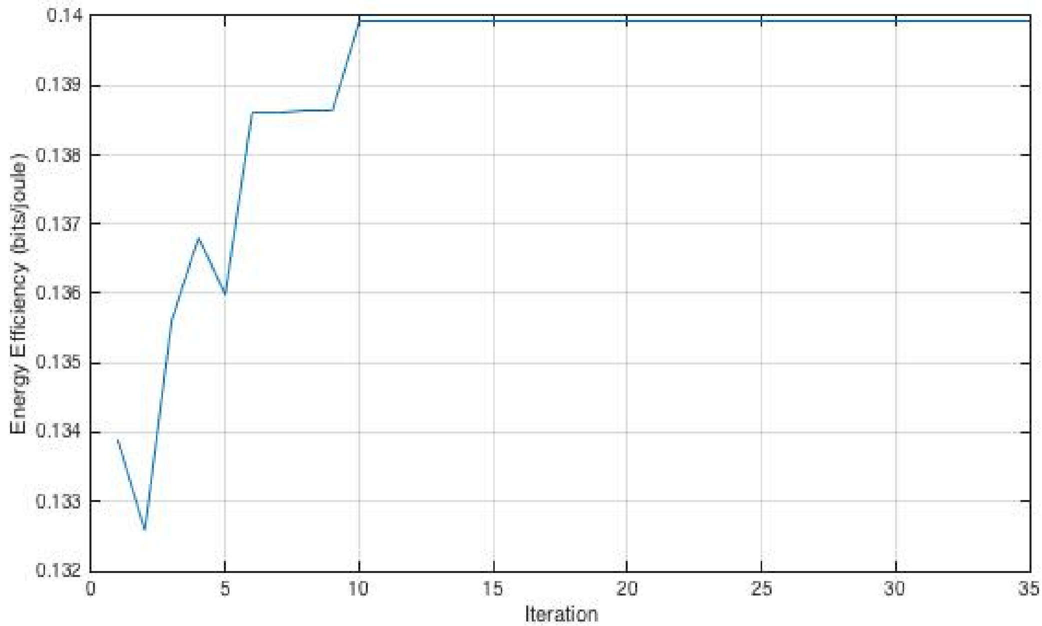


Fig. 3.19 Simulated annealing in macro cell 3 and pico cells 15-21.

setting, Table 3.2 shows the best PCs combination in each MC-PCs group, and the number in each row of the table represents the active PC's number. According to the SA, these PCs

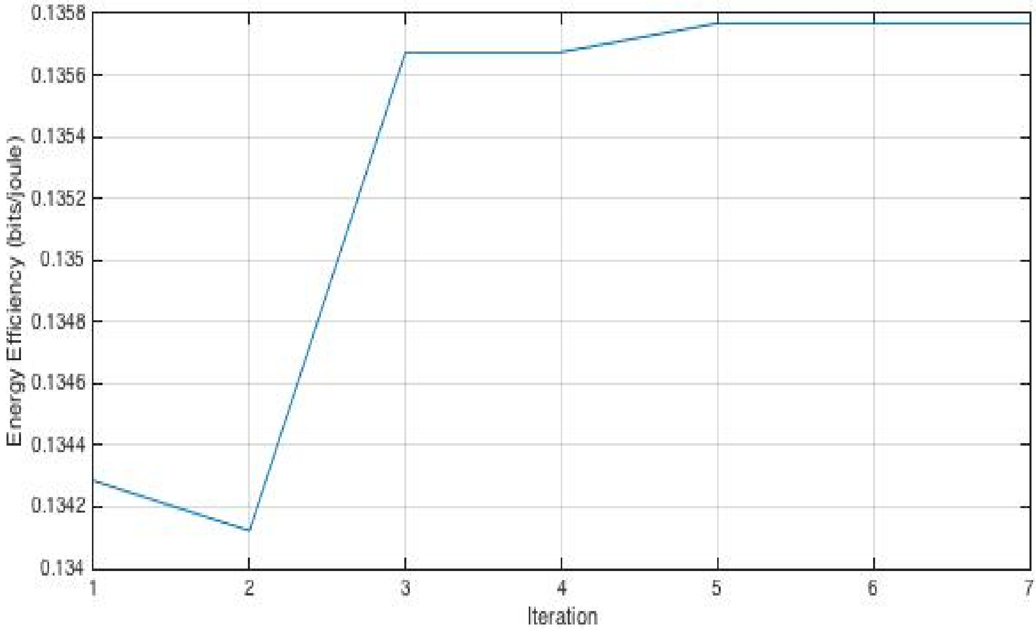


Fig. 3.20 Simulated annealing in macro cell 4 and pico cells 22-28.

Table 3.2 Pico Cells Combination

Macro Cell and Pico Cells Group	Pico Cells Combination
Macro Cell 1/ Pico Cells:1-7	1, 3, 4, 5, 6
Macro Cell 2/ Pico Cells:8-14	9, 13, 14
Macro Cell 3/ Pico Cells:15-21	15, 16, 19
Macro Cell 4/ Pico Cells:22-28	22, 23, 24, 25, 27, 28

combinations can provide the best EE to the system. The original low-load PCs are 2, 5, 8, 11, 12, 13, 16, 17, 20, 21 and 25, but the system makes the decision to switch off PCs 2 and 7 in group 1, PCs 8, 10, 11, 12 in group 2, 17, 18, 20, 21 in group 3 and 24 group 4. It can be seen that the switch-off PCs is different with the original low-load PCs. The PCs combination is made based on the PCs and subscribers location, it will not only consider the load of the PCs.

Take group 1 as an example, the original active PCs location and the location of the best PCs combination in the first MC-PCs group are shown in Figures 3.21 - 3.28. The PCs 2, 5,

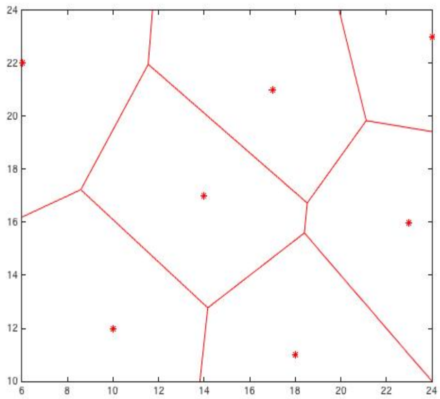


Fig. 3.21 Original pico cells group 1.

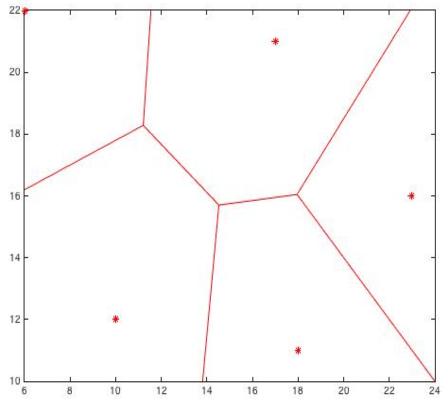


Fig. 3.22 Group 1 after simulated annealing.

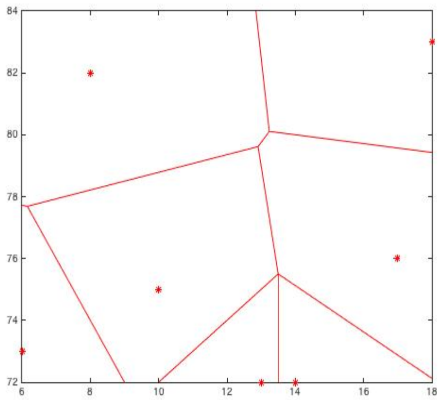


Fig. 3.23 Original pico cells group 2.

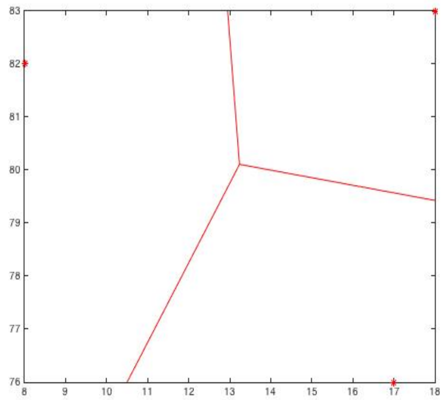


Fig. 3.24 Group 2 after simulated annealing.

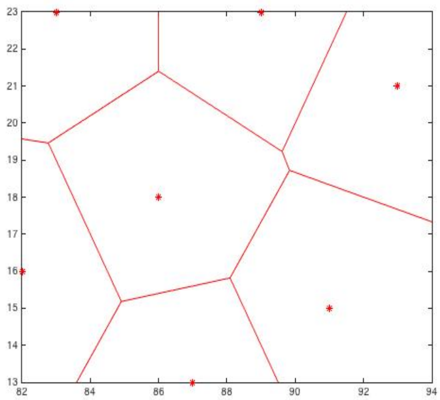


Fig. 3.25 Original pico cells group 3.

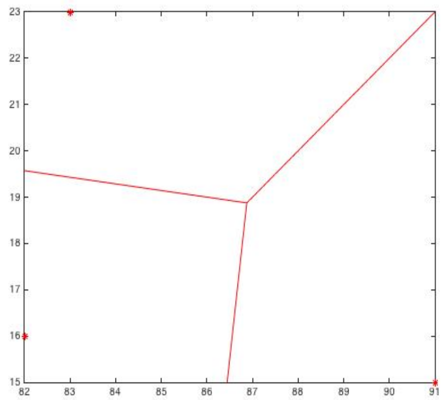


Fig. 3.26 Group 3 after simulated annealing.

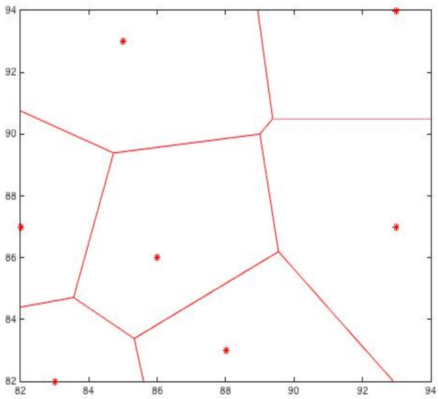


Fig. 3.27 Original pico cells group 4.

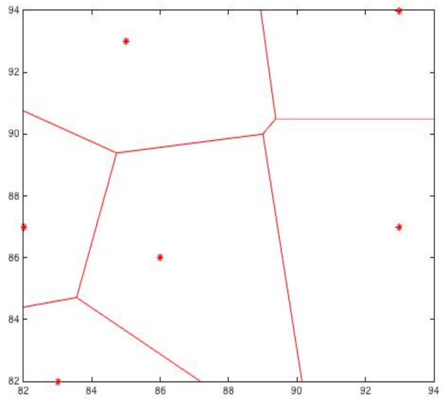


Fig. 3.28 Group 4 after simulated annealing.

7 are closed to each other, and the location of PC 5 is in the middle of PCs 2 and 7. If the system switches off the original low-load PCs 2 and 5, the subscribers within PC 2 coverage need to be connected to the remote PC 7, the subscribers will suffer from the high path loss. The PC 7 needs to allocated more resources to the subscribers to satisfy the subscribers' downlink speed. Therefore, instead of switching off original low-load PCs, the system will let PC 5 stay in active mode and switch off PCs 2 and 7, and the traffic from PCs 2 and 7 will be transferred to PC 5 after PCs 2 and 7 are switched off. It can be seen that even PCs are switched off for energy saving, it will not cause any influence to the subscribers, and the system can use minimum resources to serve the subscribers. Meanwhile, the location of PCs could become a important benefit because one PC might cover a larger area when surrounding PCs are switched off. If the PC is located in the central of all subscribers, the path loss will not be a high value, and PC can use minimum resources to serve the subscribers.

3.5.4 Resources Utilisation and Energy Efficiency

Figure 3.29 shows load of each PC after the system switches off some light-load PCs. The original active PCs number is 28, which is shown in Figure 3.16. After SA, the system will switch off 11 light-load PCs for energy saving. Compared with Figure 3.16, Figure 3.29 shows that the remaining PCs' load is increased because the load from the switch-off PCs is

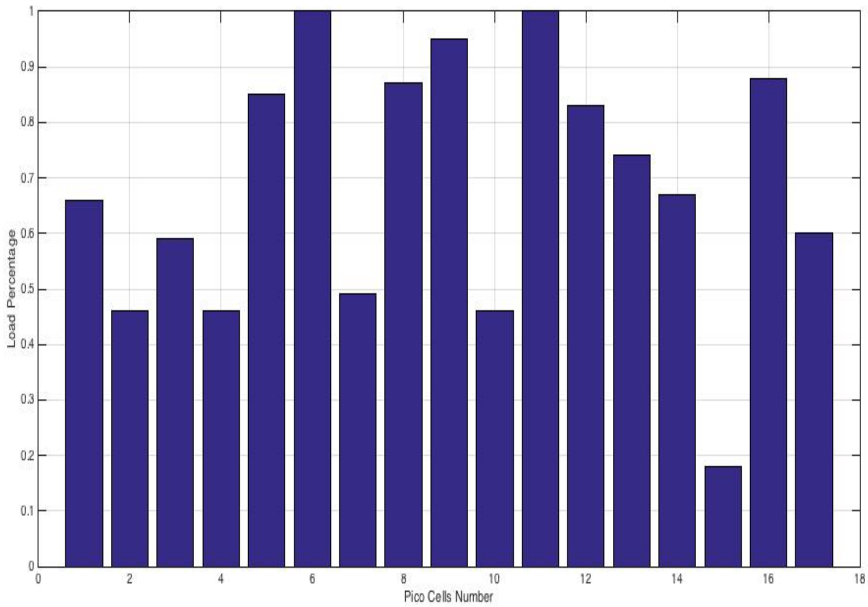


Fig. 3.29 Load of pico cells after simulated annealing.

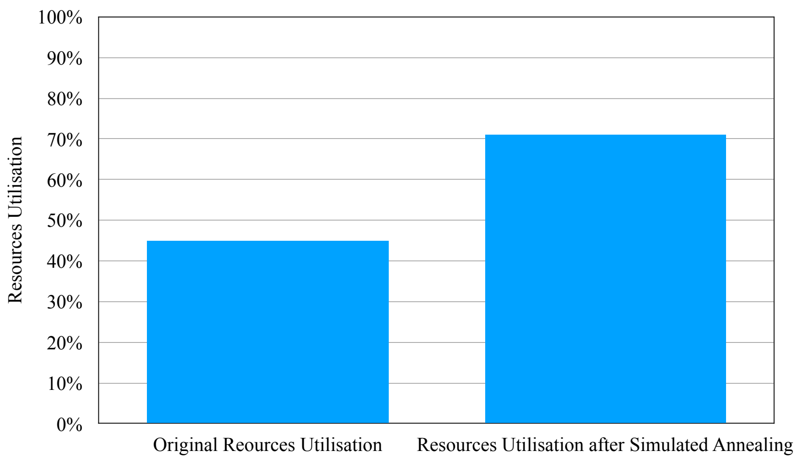


Fig. 3.30 Resources utilisation of pico cells.

redistributed to the remaining PCs. The distance between subscribers and the remaining PCs will be changed after some of the PCs are switched off. Meanwhile, the air interface of the

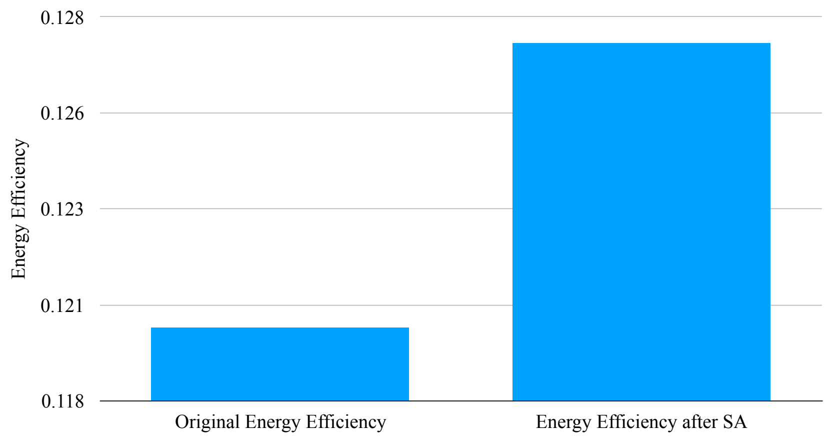


Fig. 3.31 Energy efficiency after simulated annealing.

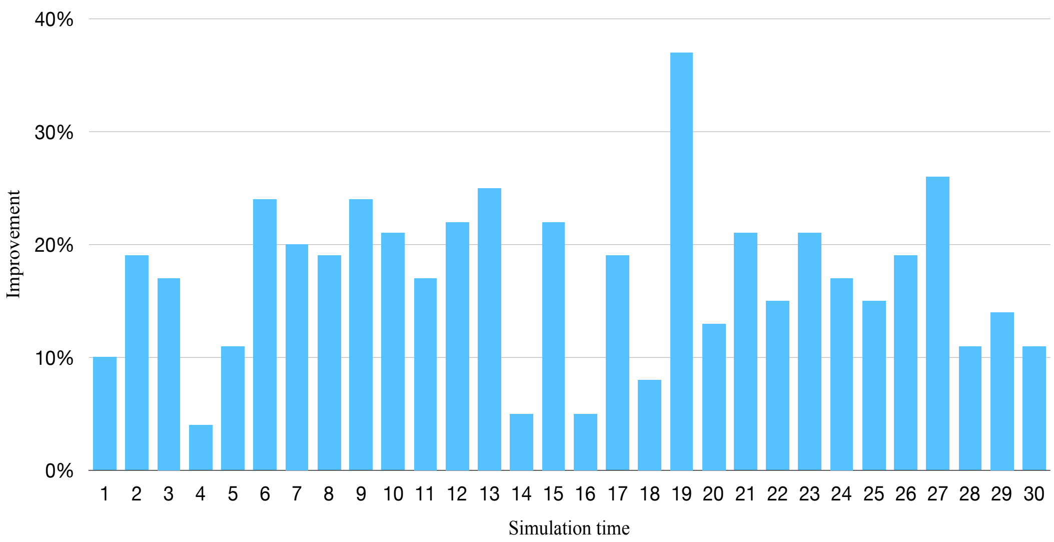


Fig. 3.32 Energy efficiency in thirty simulations.

scenario is also changed, causing the value of subscriber's SINR is different. According to Equation 3.19, if the SINR value is changed, the allocated resources will not be the same anymore. Due to the closest PCs are switched off, the SINR value will less than before, and it will force PCs to utilise more resources to serve same amount of the subscribers. It means more free BS's resources will be used to serve the existing subscribers, causing the higher resources utilisation in Figure 3.30. Therefore, switch-off strategy will improve the resources utilisation, the resource utilisation before and after switched-off strategy is 44% and 71%. The improvement is significantly.

Equation 3.15 shows the definition of the EE, the EE can be improved by wither improving the total throughput of the network or reducing the total energy consumption. Due to the total number of subscribers and required downlink are fixed in the scenario, the only way of improving the EE is by reducing the total energy consumption of the network. Figure 3.29 shows that the total active PCs number is reduced from 28 to 17, 11 PCs' power is saved after the SA algorithm, and the energy consumption is reduced. Therefore, the EE can be improved because the total throughput is a fixed value in the numerator, and the value of denominator is less than before. The EE improvement of the network is shown in Figure 3.31, the network uses less energy to serve the existing subscribers, the original EE is 0.119, and after some the light-load PCs are switched off the EE achieve 0.1273 , improving 6% EE.

Figure 3.32 shows the EE improvement in thirty simulations, the minimum EE improvement it can obtain is 4% among thirty simulations, the highest EE improvement is 37%. The EE improvement difference between every simulation is because the subscribers and the PCs will be redistributed in each simulation. The traffic distribution will not be identical. In the minimum EE improvement simulation, the small value of EE improvement is mainly because the subscribers are placed far away from the PCs, and more PCs shall be operated to serve the subscribers. However, the PCs switched-off strategy via SA algorithm can still improve the EE of the network, and the average EE improvement in these thirty simulation is 17.06%. The above discussion shows the way of measuring EE, given the condition of total number of subscribers and required downlink are fixed, the meaning of average 17.06% EE improvement means that average 17% of the total number of PCs can be switched off, and maximum 37% of the total number of PCs can be switched off among thirty simulations, one third of the PCs are switched off without reducing the overall network performance.

3.6 Conclusions

In this chapter, low-load PCs are switched off to reduce the energy consumption, and SA algorithm is proposed to choose the best PCs combination, aiming to achieve optimal EE. The optimisation problem is considered as a combination problem that the system going to select a certain number of PC from the PC group to serve the current subscriber. However, the optimisation problem will become a NP-hard problem with large number of PC and subscriber. Therefore, sub-optimisation problem is formulated to reduce the computation complexity, considering the scenario that a set of PCs are deployed within MC coverage.

In SA algorithm, the cost function value will be estimated when system choose a new PCs combination, and the algorithm will accept the new PCs combination if the value of cost function is larger than the previous PCs combination. Moreover, algorithm will accept worst PCs combination if the criterion is satisfied, aiming to avoid the local maximum valuer, but the accepting chance will be reduced when the iteration is increased. After all PCs combinations are tested, the best PCs combination can be obtained.

The simulation result gives the conclusion that the system might not achieve the best EE if the original low-load PCs are switched off. Because the subscribers might connect to other remote PC after the low-load BSs are switched off, and the system shall allocate more resources to provide sufficient service. Due to the chosen PCs' locations are different in each PCs combination, the subscriber's SINR might not be identical in each PCs combination. On the aspect of resources utilisation, the traffic is redistributed to neighbour PCs after low-load PCs are switched off. The simulation also shows that the EE can be improved when resources utilisation is increased given the same number of PCs are switched off. From the aspect of energy saving, compared with full power PCs, the energy consumption of network can be reduced.

Chapter 4

Energy Efficiency Enhancement via Naive Bayesian Classifier Traffic Prediction

4.1 Introduction

In the last two decades, mobile phone becomes an affordable product to the public, and advanced wireless technology creates a new way for information exchange. However, the original BS deployment cannot match the growing traffic demand, and a new type of low transmission power BS, PC is deployed within MC coverage to provide extra access point [82]. This new type network architecture will become a complicated network system because the complicated air interface. Moreover, the increasing number of BS will consume considerable energy, and the energy consumption problem becomes a top issue to the network operators. However, the traffic demand in the urban city is changed frequently, especially in high subscribers density area. Therefore, PC switch-off strategy in light-traffic period is proposed in the previous chapter to improve the EE and resources utilisation. Traffic demand is similar to the car traffic that it is growing proportional with the increasing number, and the growing traffic will bring many new challenges to network operators, such as traffic congestion, power consumption etc [102]. A small number of PCs cannot provide sufficient capacity if traffic is increased, the system should activate more PCs to provide regular coverage and service [103]. Therefore, an accurate traffic prediction result should be obtained to let the system informs the switch-off PCs return to regular transmission power to serve the increasing subscribers.

The traffic prediction in CN will be more difficult when subscriber is moving randomly. Traditional method, Markov model is utilised to predict the traffic when the subscriber is moving within the network [50]. However, it only calculates the transition probability from one BS to the remaining BSs, and the subscriber information is not consider in the prediction. If the system needs a reliable prediction result, subscriber information should be considered. In [104], Bayesian network is applied to predict the car traffic, the probability of each state is calculated, and the prediction result will be made based on the probability of the state. The NBC is a probabilistic model, and it is widely used in many other areas. In [105], NBC is proposed to discover the network failure by analysing user data. The collected user data will help the system to detect the failure within the network. NBC is utilised to classify load of BS in [106], the system can determine the low-load BS by different BS's feature, and it will switch off low-load PCs for energy saving. The Bayesian model is applied in [107] to predict the mobility based on the occurrence of the link duration before break between two nodes. The parameters of average encounter rate and node degree are collected to predict the velocity of node [108].

Traffic pattern is complicated and vary significantly in real CN, depending on time and the location of BS [109]. Inspired by [105], [106], [107] and [108], the subscriber's information can be collected and utilised for traffic prediction. Moreover, traffic could be predictable due to the subscriber might have a fixed moving pattern and behaviour around their working place. Based on the data collection, analysis and learning, subscriber's features can be utilised to determine the traffic pattern in specific timing. The system can know the incoming traffic with specific timing, and it can appropriate modify PC's transmission power to prepare the incoming traffic. New techniques, like ML technics are proposed to utilise in CN [110]. NBC is utilised in this chapter to predict subscriber's destination via serval collected subscriber's features. According to the probability of each BS, NBC can provide the prediction result when a new subscriber data is collected. The result will be utilised for BS transmission power modification, PC's transmission power is increased proportional to the growing traffic, increasing the coverage of PC. The simulation will show the prediction accuracy via NBC and EE improvement after the transmission power is modified.

The remaining parts are organised as the following structure. Section two will describe the system model, including the notation, description of scenario and the system process. NBC model will be introduced in Section three, including the derivation. Data analysis and traffic prediction via NBC will be shown in this section. Section four will show the traffic prediction result. Section five will introduce the power modification model, and this section will show the energy consumption and EE after the power of PC is modified. The last section will draw the conclusion.

4.2 System Model

4.2.1 Notation

Let's consider a HetNet with two BS tiers, MC tier and PC tier, a set of PCs are deployed within MC coverage, and let's denote a set of PCs P and a single MC M, where $\{1, 2, 3, ...p\} \in P$. Let's define the subscriber group, $\{O_1, O_2, O_3, ...O_i\}$, and each subscriber O_i will contain up to d subscriber features, where $\{o_1, o_2, o_3, ...o_d\} \in O_i$.

4.2.2 Scenario

The scenario will begin with a light traffic demand, the MC and PCs deployment is shown in Figure 4.1. Two horizontal axes represent x and y value in the figure, and the value represent the pixel. In this figure, one pixel will represent 10 meters. Meanwhile, the vertical axis represents the signal strength from nearby base stations, it can be represented by dBm. The MC will provide a basic coverage, and PCs are deployed in the high traffic demand area during the working hours.

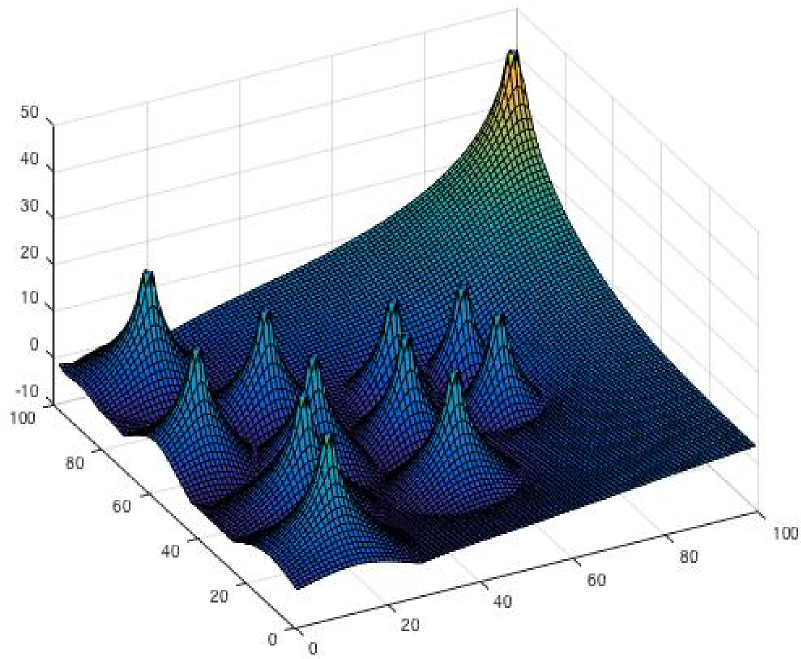


Fig. 4.1 Scenario.

11 PCs are deployed within MC coverage, the locations of the PCs are the highest points in Figure 4.1, the yellow one on the top right corner is the location of the MC. Meanwhile, it is assumed that PCs 6 and 8 have respective metro exits in their coverage, which is shown in Figure 4.2 as a grey colour square. In this scenario, subscriber will first appear in either one of the metro exits, and it is assumed that half subscribers will appear in PC 6, and the remaining subscribers will appear in PC 8. Except the PCs 6 and 8, the remaining PCs are considered as subscriber's destination, and the probability of subscriber reaches each PC is assumed to be identical. During the subscriber movement, it will be served by either PC or MC, and the serving BS is determined by the strongest received signal strength from PCs and MC. It is assumed that the starting timing is 8:45, and each subscriber will appear in

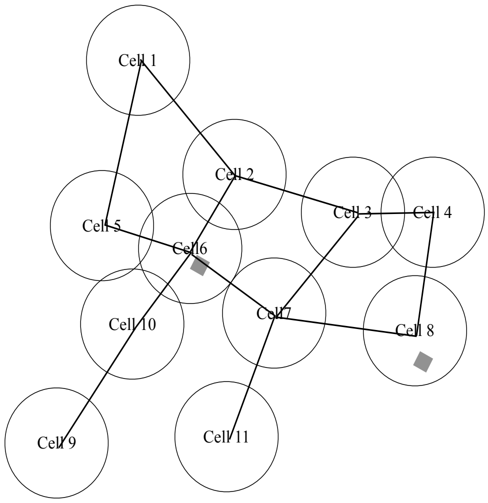


Fig. 4.2 Routes.

these two metro exits with different timing, and the ending time is set 15 minutes after the beginning timing. In each time slot, same amount of subscriber will appear in the two exits, and all subscribers attempt to reach the destination before 9:00.

Table 4.1 Pass pico cells between starting pico cell and ending pico cell.

Pass Pico Cells		
Starting pico cell	Ending pico cell	Pass pico cell number
6	1	6-5-1/6-2-1
6	2	6-2
6	3	6-2-3/6-7-3
6	4	6-2-3-4/6-7-3-4
6	5	6-5
6	7	6-7
6	9	6-10-9
6	10	6-10
6	11	6-7-11
8	3	8-4-3/8-7-3
8	4	8-4
8	7	8-7
8	11	8-7-11

It is assumed that the subscriber will have the following moving pattern. No matter if the remaining time is insufficient or not, subscriber will walk to the destination with high speed. If the remaining time is insufficient, subscriber will walk slowly to close PCs group. The subscriber will not have any specific moving pattern if the destination is in close PC group and remaining time is sufficient. The pass PCs between metro exits and destination are shown in Table 4.1, and routes are shown in Figure 4.2 as solid line. In this scenario, the subscribers requires different level of downlink service. During the walking, it is assumed that the downlink requirement is inversely proportional to the walking speed.

4.2.3 System Process

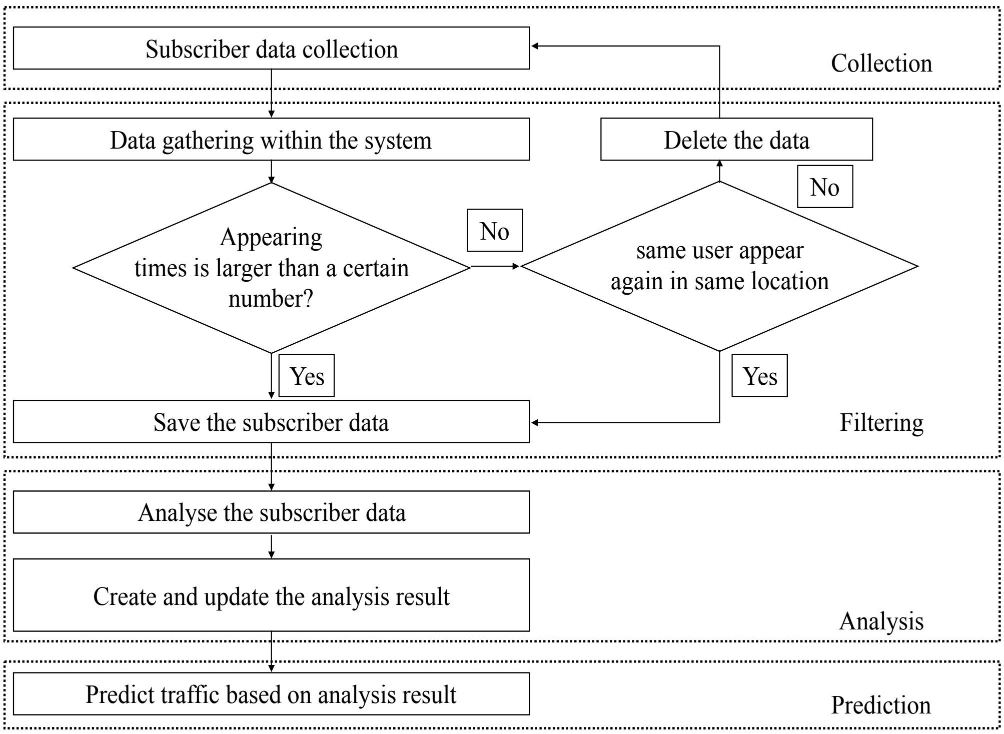


Fig. 4.3 User data base Naive Bayesian classifier prediction process.

The system process will include four parts, data collection, filtering, analysis, and prediction, which are shown in Figure 4.3.

Collection The main function of this phase is to collect the subscriber's data when they connect to the network. According to the total number of connected BS, it can create a data matrix to store the subscriber's data for analysis. The size of the data matrix will be $(n+m)$ $\times$ d, $n+m$ represents the connected BS, and d represents the total number of subscriber's feature. The historical data matrix of subscriber i in the network from PC p_n to p_{n+m} can be represented by the following matrix.

$$H_{n,m} = \left\{ \begin{array}{ccccc} h_1^n, & h_2^n, & h_3^n, & \ldots & h_d^n \\ h_1^{n+1}, & h_2^{n+1}, & h_3^{n+1}, & \ldots & h_d^{n+1} \\ h_1^{n+2}, & h_2^{n+2}, & h_3^{n+2}, & \ldots & h_d^{n+2} \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ h_1^{n+m}, & h_2^{n+m}, & h_3^{n+m}, & \ldots & h_d^{n+m} \end{array} \right\} \tag{4.1}$$

where h_d^n represents the subscriber i is served by PC p_n, and the value of feature d is collected by the PC. $H_{n,m}$ represents the daily movement from PC n to PC $n+m$. If the subscriber appears again in the network, this subscriber's feature will be recorded. Therefore, it can be seen that there will be five sets of subscriber data during the working day, and traffic prediction will be made based on this collected data. The subscriber's features will be walking speed, downlink, remaining time and direction in this chapter.

Filtering Commuters who work in a specific place are the ideal subscribers' data for analysis and traffic prediction. The routine movement and appearing timing can be considered as a fixed pattern, such as fixed paths from one place to the other, and commuters also have a pattern on time domain because they will frequently appear in a certain timing, such as the time around working hours. The filtering part is proposed to prevent the system utilises temporary subscriber's data to determine the pattern and use for the traffic prediction. The data of tourist will not suitable for traffic prediction because most of them will only appear in the same location a few times, and this data cannot construct a specific traffic pattern. The temporary data might reduce the prediction accuracy. Therefore, when subscriber's data is collected by the system, part of the data will be filtered before it moves to the next phase. However, data will not be deleted immediately, and it will be saved for a short period to prevent some new subscribers might not have record before. If new subscriber frequently appears in a specific location, the system will also transfer this subscriber data into data analysis phase and uses it for traffic prediction.

Analysis and Prediction The data analysis phase will analyse the subscriber's data and determine the characteristic of feature, and the analysis result will be utilised for the traffic prediction. The data analysis phase can be considered as the preparation of prediction. The detail of analysis and prediction phases are shown in the following section.

4.3 Traffic Prediction via Naive Bayesian Classifier

4.3.1 Naive Bayesian Classifier

The NBC is a probabilistic classifier with a high computational efficiency, based on the previous collected data, the classification result can be obtained, and NBC is widely utilised for the prediction. The advantage of using NBC is this classifier does not need too much training data [111], the data issue will not be a main problem when subscribers number in urban city is huge. However, the subscribers in remote area is little, it will take a long period to construct a reliable prediction model to predict the user destination. By applying NBC, the predictor can work even the amount of data is limited. Moreover, the NBC can use both continues and discrete data to predict the subscriber's destination, the features are chosen depending on the reliability of providing accurate subscriber's destination. Therefore, it can not promise that the features will be only continues data or discrete data. Therefore, NBC will become a ideal predictor that it can use both data type. Meanwhile, compared to the K-Nearest Neighbour classifier (KNN), NBC is a fast model to predict the subscriber's destination [112], and it can be used for real-time prediction. It is important to obtain the subscriber's destination within a short period because some of the subscribers will move quick within the network, such as the subscribers in metro or railway station. The NBC has such ability to provide the prediction result instantly, and the network system can react one step forward than the subscribers to prepare the incoming traffic.Therefore, the above advantages can greatly help the traffic prediction in CN.

The classifier is based on Bayes theorem and constructed on the assumption that features are mutually independence [41] [108] [113]. Based on the scenario, subscriber O_i contains d mutually independence features. Therefore, the object can be represented by the following relationship.

$$O_i = (o_1, o_2, o_3, ..., o_d) \tag{4.2}$$

In this chapter, the objective is used to determine the subscriber i reaches one PC's probability given a set of collected features value $(o_1, o_2, o_3, ..., o_d)$. Let's define the prediction result as R_p, which is the probability of subscriber reaching one PC's probability. Therefore, if p PCs are deployed in the scenario, system will provide the same number of probability outcomes $p(R_p \mid O_i)$. According to the Bayes theorem, the probability can be represented by Equation 4.3.

$$p(R_p \mid O_i) = \frac{p(R_p \cap O_i)}{p(O_i)} \tag{4.3}$$

Another equation can be obtained.

$$p(O_i \mid R_p) = \frac{p(O_i \cap R_p)}{p(R_p)} \tag{4.4}$$

Therefore, according to Equation 4.4, the following equation can be obtained.

$$p(O_i \cap R_p) = p(O_i \mid R_p)p(R_p) \tag{4.5}$$

Because $p(R_p \cap O_i) = p(O_i \cap R_p)$, let's combine the Equations 4.3 and 4.5, it can obtain the following equation

$$p(R_p \mid O_i) = \frac{p(O_i \mid R_p)p(R_p)}{p(O_i)} \tag{4.6}$$

where $p(R_p \mid O_i)$ is called posterior probability, $p(O_i \mid R_p)$ is called likelihood, $p(R_p)$ is called prior probability and $p(O_i)$ is called evidence.

The object O_i contains d features, and the subscriber's features value can be calculated based to the observation, and the probability of each feature can be represented as $p(o_1 \mid R_p)$, $p(o_1 \mid R_p)$,... $p(o_d \mid R_p)$. Therefore, the probability of $p(O_i \mid R_p)$ can be obtained. $p(R_p)$ represents the probability of subscriber reaching one PC. Based on the assumption that the subscriber will have the same probability to reach each PC, it means that $p(R_1) = p(R_2) = p(R_3) = ...p(R_p)$. The denominator $p(O_i)$ can be represented as

$$p(O_i) = \sum_{1}^{p} p(R_p)p(O_i \mid R_p) \tag{4.7}$$

Because the denominator does not depend on the R_p when the features o_d are given, and denominator can be considered as a constant. Therefore, the probability $p(R_p \mid O_i)$ only needs to consider the numerator.

Let's consider the joint probability $p(O_i \mid R_p)$ in Equation 4.3, and it can be represented as

$$p(o_1, o_2, o_3, ..., o_d, R_p) = p(o_1 \mid o_2, ..., o_d, R_p)p(o_2, o_3, ..., o_d, R_p) \tag{4.8}$$

$$p(\bigcap_{k=1}^{n} A_k) = \prod_{k=1}^{n} p(A_k \mid \bigcap_{j=1}^{k-1} A_j) \tag{4.9}$$

The chain rule is shown in Equation 4.9, and let's apply the chain rule in Equation 4.8, the numerator can be represented by the following equation

$$\begin{aligned} p(o_1, o_2, o_3, ..., o_d, R_p) &= p(o_1 \mid o_2, ..., o_d, R_p) p(o_2, o_3, ..., o_d, R_p) \\ &= p(o_1 \mid o_2, ..., o_d, R_p) p(o_2 \mid o_3, ..., o_d, R_p) \\ &\quad p(o_3, ..., o_d, R_p) \\ &= p(o_1 \mid o_2, ..., o_d, R_p) p(o_2 \mid o_3, ..., o_d, R_p) \\ &\quad ...p(o_{d-1} \mid o_d, R_p) p(o_d, R_p) p(R_p) \end{aligned} \tag{4.10}$$

Based on the assumption of each feature is mutually independent, one feature o_i will only realtive to the probability R_p, it means that the remaining feature can not cause any influence to the R_p. Therefore, and the following relationship can be obtained

$$p(o_1 \mid o_2, o_3, ..., o_d, R_p) = p(o_1 \mid R_p) \tag{4.11}$$

Therefore, the numerator in Equation 4.10 can be represented as

$$p(O_i \mid R_p) = \prod_{1}^{d} p(o_d \mid R_p) \tag{4.12}$$

Let's combine Equations 4.7, 4.12, the following equation can be obtained that it represents the observed object reach one PC's probability.

$$p(R_p \mid O_i) = \frac{p(R_p) \prod_1^d p(o_d \mid R_p)}{\sum_1^p p(R_p) p(O_i \mid R_p)} \tag{4.13}$$

As it mentioned before, the denominator will be considered as a constant if the value of the feature is known, it means that only the numerator will influence the value of $p(R_p \mid O_i)$. Therefore, the objective of classifier is to determine which PC has the highest probability

$p(R_p \mid O_i)$ given a set of features value $(o_1, o_2, o_3, ..., o_d)$ from subscriber O_i. If the system contains p PCs, classifier can provide the same number of probability. The prediction problem can be represented by the following expression

$$R_p = \underset{\{1,2,3...,p\}}{\arg\max}\ p(R_p) \prod_{1}^{d} p(o_d \mid R_p) \tag{4.14}$$

When subscriber's features are collected by the system, it will generate corresponding PC's probability based on the previous collected data. If subscriber's features value is similar to the previous collected data, the probability will be a large value. NBC will choose the highest PC's probability as the prediction result, and it represents the subscriber might move to this PC.

4.3.2 Features Value Analysis

After the temporary subscriber's data is filtered by the system, the remaining data will be transferred to analysis phase, and the analysis result is stored in the system. The features will be categorised into two parts, continues features and discrete features in this chapter.

Continues Feature In this chapter, the features of walking speed, downlink and remaining timing are considered as continues features. When subscriber appears in either PC 6 or 8, the subscriber will connect to the starting PC first, and the subscriber features value will be recorded. Therefore, the probability of these features can be calculated. The collected subscriber features value between PC 6 and PC 1 is shown in Table 4.2.

After the system collects the subscriber features values, it will calculate the mean μ and variance σ^2. Based on the scenario, starting PCs are 6 and 8, and subscriber will move to the remaining PCs. Therefore, $2 \times p$ PCs combination can be obtained. The Table 4.3 can be obtained after the features are analysed by the system, the table shows the mean and variance of each feature in each PCs pair.

Table 4.2 Subscriber features between pico cell 6 and 1.

Subscriber information			
Subscriber No.	Speed (m/s)	Downlink (dps/s)	Remaining time (mins)
8	1.79	86	8:52
11	2	64	8:45
15	1.70	86	8:47
16	2.04	64	8:46
27	1.75	86	8:59
31	1.5	102	8:57
38	1.5	111	8:56
43	1.3	126	8:58
...	...	...	...
...	...	...	...
989	1.4	121	8:50
994	2	64	8:49
998	2	64	8:49

Table 4.3 Mean and variance of subscriber features value in each pico cells combination.

Collected subscriber features value						
Pico Cell	Mean(speed)	Variance(speed)	Mean(downlink)	Variance(downlink)	Mean(time)	Variance(time)
6-1	1.68	0.04	94.03	0.452xe^3	8.7	15.9
6-2	0.74	0.02	240.8	1.176xe^3	8.6	15.6
6-3	1.27	0.02	141.7	0.5xe^3	11.7	3.7
6-4	1.28	0.01	140.9	0.278xe^3	11.6	3.6
6-5	0.8	0.01	215.43	0.9xe^3	7.3	19
6-7	0.7	0.02	241.88	1.42xe^3	7.47	19.1
6-9	1.24	0.01	144.8	0.398xe^3	5.5	11.3
6-10	0.76	0.02	236.2	1.44xe^3	8.1	20.9
6-11	1.21	0.01	149.3	0.35xe^3	11.88	6.32
8-3	1.25	0.02	143.9	0.46xe^3	6.73	16.6
8-4	0.74	0.02	240.1	1.19xe^3	7.6	16.8
8-7	0.7	0.01	241.9	1.07xe^3	6.8	17.7
8-11	1.65	0.06	99.6	0.756xe^3	8.8	16.4

Discrete Feature The subscriber's moving direction is considered as discrete feature value in this chapter. It is assumed that there are four directions, which are north, east, south and west. According to the subscriber's moving direction, the direction indicator will assign the direction to subscriber after they connect to the starting PC. Therefore, subscriber has only one direction when subscriber chooses the path to their destination. Let's define the D_{ij}^{s-e} as the direction of subscriber O_i when subscriber moves from PC s to PC e with the direction j. The direction value will equal to 1 if subscriber moves to their destination with the corresponding direction, 0 otherwise. $\sum_{i \in I} P_i = 1$ Therefore, probability of direction feature o_j can be calculated. According to the Equation 4.15, the probability of direction feature is shown in Table 4.4.

$$p(o_j \mid R_e) = \frac{\sum_{i=1}^{I} D_{ij}^{s-e}}{\sum_{i=1}^{I} \sum_{j=1}^{J} D_{ij}^{s-e}}, \; where \; \forall j \in J \tag{4.15}$$

Table 4.4 Direction Feature Probability.

	Direction			
Base station	North	East	South	West
6-1	0.458	0	0	0.541
6-2	1	0	0	0
6-3	0.7	0.3	0	0
6-4	0.42	0.57	0	0
6-5	0	0	0	1
6-7	0	1	0	0
6-9	0	0	1	0
6-10	0	0	1	0
6-11	0	1	0	0
8-3	0.5	0	0.5	0
8-4	1	0	0	0
8-7	0	0	1	0
8-11	0	0	1	0

After the data analysis, these probability distributions will be utilised for the traffic prediction. Meanwhile, the system will continue to collect data for updating. By doing this, the influence from the changing environment can be reduced, improving the prediction accuracy. Moreover, if any subscriber appears in a certain location larger than a threshold, this subscriber data will be put into analysis.

4.3.3 Traffic Prediction

The traffic prediction will be made based on the collected subscriber's data. After the system collects a subscriber data from the network, it will determine which PC parameters can match the new subscriber data, and it will provide the prediction that the predicted PC has the highest probability. It is assumed that these features are followed Gaussian distribution and mutually independent. Therefore, the probability distribution of subscriber feature $p(o_d \mid R_p)$ given the feature o_d can be calculated by Equation 4.16.

$$p(o_d \mid R_p) = \frac{1}{\sqrt{2\pi\sigma_d^2}} exp(\frac{-(o_d - \mu_d)^2}{2\sigma_d^2}) \tag{4.16}$$

According to the Equation 4.16, the higher probability can be obtained if the collected subscriber feature is similar to the prior collected feature value, otherwise the feature value will be small, and it represents this feature can not match the characteristic of this PC. If one subscriber contains d features, the same number of features probability distribution $p(o_1 \mid R_p), p(o_2 \mid R_p), p(o_3 \mid R_p), ...p(o_d \mid R_p)$ can be obtained in one PC. After all features probability sum together, the classifier will choose the highest probability PC as the prediction result. Therefore, if p PCs are deployed within the network, one subscriber feature o_d will have up to p probability distribution in different PC, where $p(o_d \mid R_1)$, $p(o_d \mid R_2)$, $p(o_d \mid R_3)$,... $p(o_d \mid R_p)$. According to Equation 4.14, classifier needs to determine the value of $\prod_1^d p(o_d \mid R_p)$, and the prediction result will be made depending on which PC has the highest probability.

The feature chosen is crucial in NBC, the advantage of NBC is that many features involve the prediction. Although part of the subscriber features might not accurate or related to the previous collected data, the NBC can slightly minimise the influence because the final probability is the sum of all features probabilities. Moreover, it can be seen that if the system chooses uncorrelated features for traffic prediction, the prediction accuracy will be significantly reduced. It is because the probability of one uncorrelated feature reaching one destination will not generate a certain pattern, and the unpredictable probability will goes up and down that it will lower the combined probability in the final step of the prediction. The other downside is the increased data storage and the calculation time, the time will spend on these features that will not bring any contribution to the user destination prediction.

4.4 Simulation Result

The simulation result is shown in the following figures, including the collected subscribers data, prediction accuracy with different randomness and different features.

4.4.1 Features Value of Pico Cell

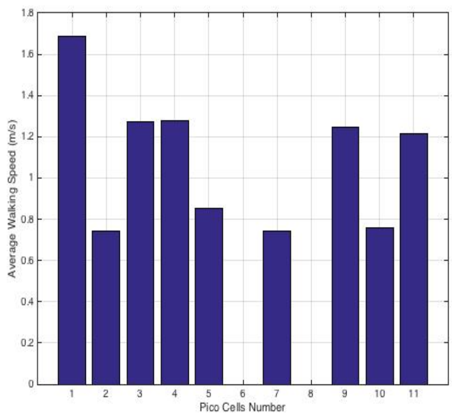


Fig. 4.4 Average walking speed in pico cell 6.

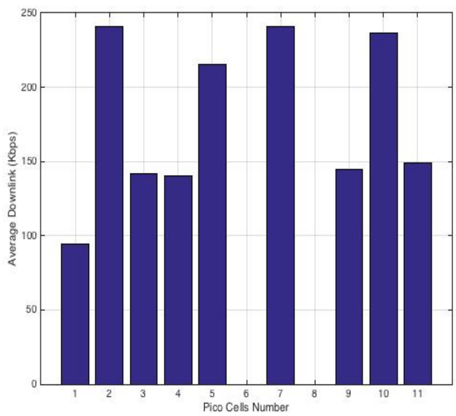


Fig. 4.5 Average downlink in pico cell 6.

Figures from 4.4 to 4.9 show the mean of walking speed and downlink and remaining time after the system collects and analyses the subscriber features, and each starting PC has its distinct characteristics. According to the figures, the PCs can be divided into two groups,

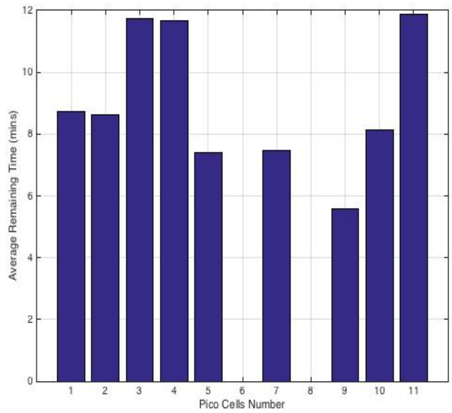


Fig. 4.6 Average remaining time in pico cell 6.

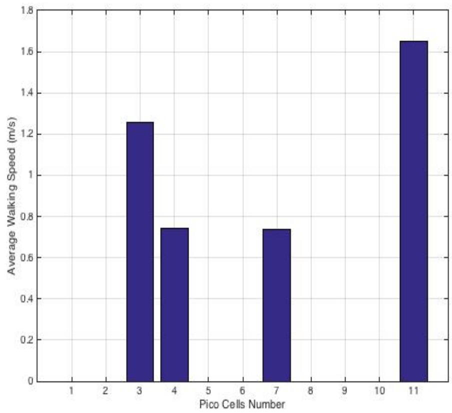


Fig. 4.7 Average walking speed in pico cell 8.

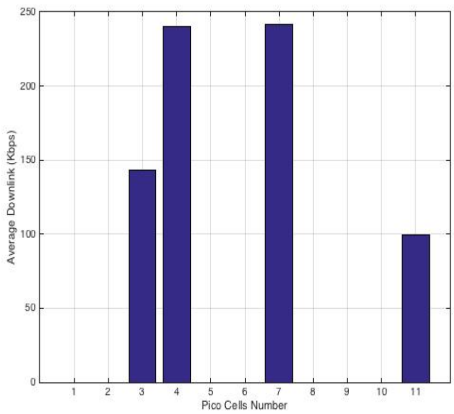


Fig. 4.8 Average downlink in pico cell 8.

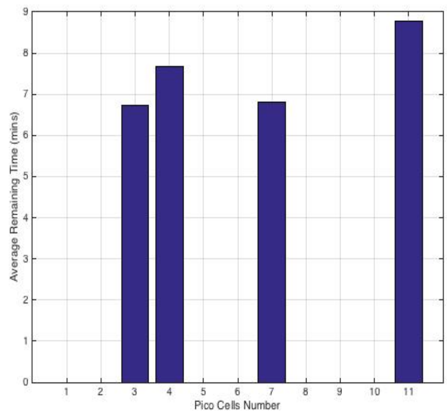


Fig. 4.9 Average remaining time in pico cell 8.

remote PC group and close PC group. Take average walking speed of PC 6 as an example, the remote PCs has faster walking speed than the close PCs, which is above 0.8 m/s in Figure 4.4. PCs 2, 5, 7 and 10 are close to starting PC 6, the average walking speed is moved up and down by 0.8 m/s. Moreover, subscriber will require less downlink when they walk with high speed, which is shown in Figures 4.5 and 4.8. According to these features value, the prediction accuracy is 65%. Due to the similar features value in some PCs, the system cannot distinguish the difference because the probability of one features is very closed to each other in serval PCs. For example, one user might reach one PC with a certain speed, but this speed

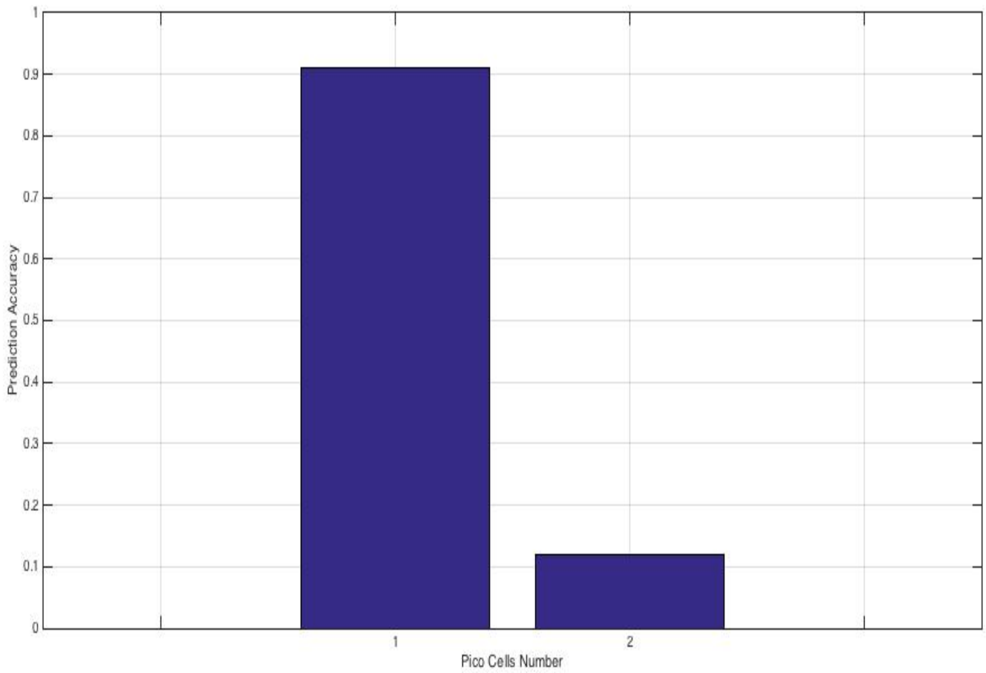


Fig. 4.10 Prediction accuracy in different distance pico cell group.

might be recorded in many other PCs as well. Take starting PC 6 as an example, the PCs 2, 5, 7 and 10 have similar features value in walking speed, which is shown in Figure 4.4. When the system calculates the feature's probability, each PC has similar value of $\prod_1^d p(o_d \mid R_p)$, and the classifier will provide the wrong prediction. The similar situation can be found in the Figure 4.6 that bars 1 and 2 have closed value. Figure 4.10 represents the prediction accuracy in different distance group, and 2 PCs are chosen, which are the PC 1 in remote PC group and PC 2 in close PC group. According to the figure, it can be seen that the prediction accuracy in remote distance group is much higher than close distance group, 91% prediction accuracy in PC 1 and 12% in PC 2. It can be seen that the selected features cannot distinguish the difference in the close PC group. Meanwhile, the features value in remote PC group is more distinct than close PC group, the system will have fewer chance to make the wrong prediction. Therefore, a well-chosen feature should has ability to distinguish the difference among similar PCs.

4.4.2 Prediction with Direction Feature

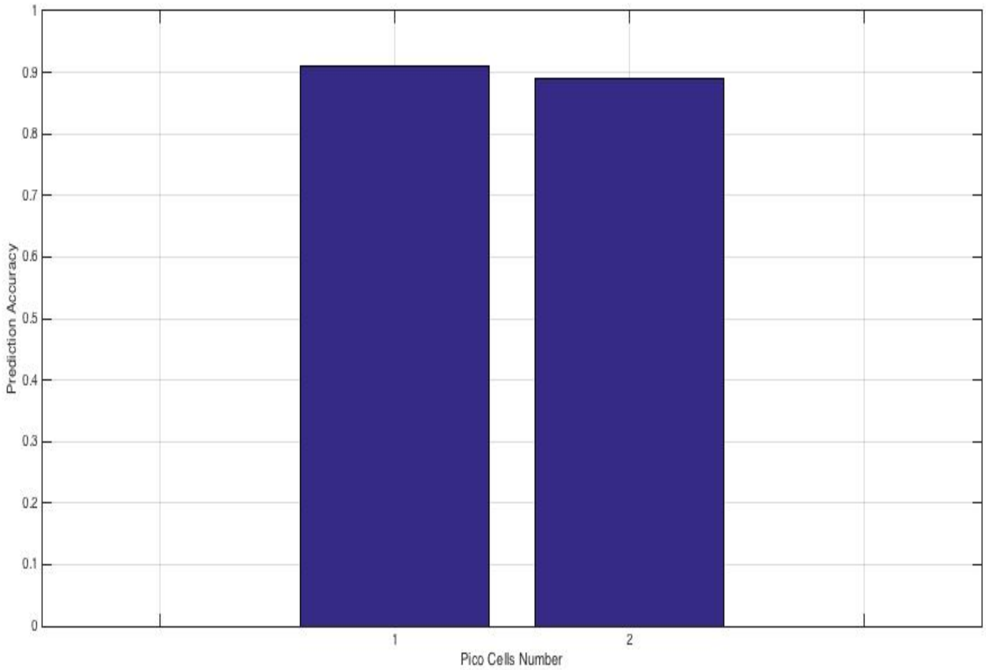


Fig. 4.11 Prediction accuracy improvement with direction feature.

Due to the system cannot distinguish the difference in close PC group, the following section will bring the direction feature into traffic prediction. Table 4.4 shows the probability of subscriber's direction feature after the system analyses the collected data. After the direction feature is added into the prediction, the overall prediction accuracy is increased to 92%. The Figure 4.11 shows the prediction accuracy comparison between the previous PCs, PC 1 and PC 2. The figure shows that PC 2 prediction accuracy is increased to 89%, 77% improvement. Because the direction feature can let the system distinguishes the similar PCs in close PC group. After the system determines the subscriber will move to the close PC group, the direction feature can let the system further predict the subscriber movement. It can be seen that the direction feature build a strong connection between PCs and feature value, and it helps the system to distinguish the difference in close PCs group.

However, if the classifier only utilises direction feature to predict subscriber destination, the prediction accuracy has only 38%. Although direction feature bring a great contribution in prediction accuracy increment, but if the prediction is only made by direction feature, the feature cannot tell the difference when some PCs are located in the same direction. Take starting PC 6 as an example, the PCs 1, 2, 3 and 4 are all on the north side, and classifier will contain 4 possible destinations if the subscriber's direction feature shows the north direction, but direction feature can not tell the difference among these 4 PCs.

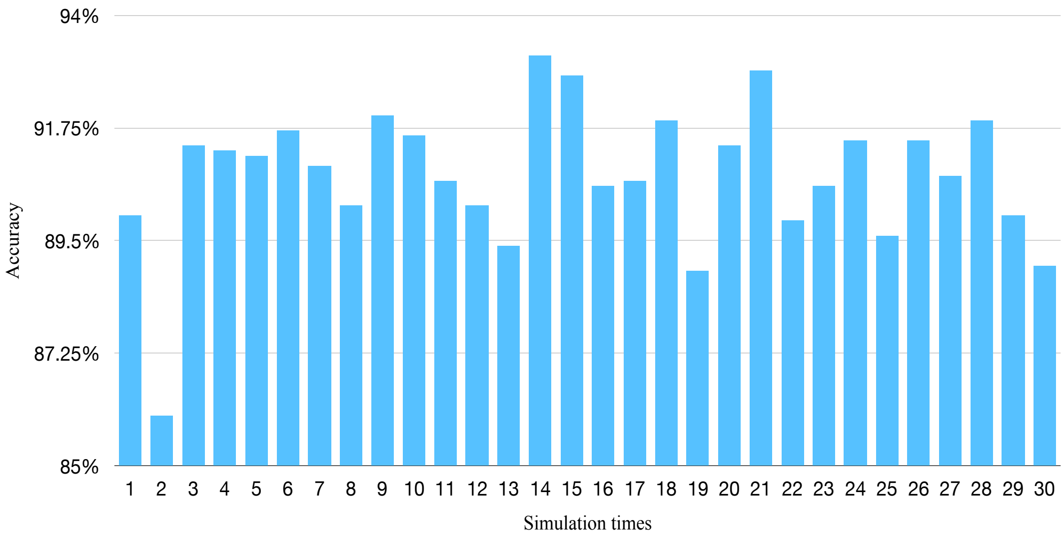


Fig. 4.12 Prediction accuracy of thirty scenarios.

The NBC is used in multiple scenarios to test the prediction accuracy, NBC is applied in thirty different scenarios and the prediction accuracy is shown in Figure 4.12. From the figure, it can be seen that the worst and the best prediction accuracy can be obtained are 86% and 93.2%. The average prediction accuracy is 90.8%. The multiple simulations can clearly show that the prediction can be significantly improved and stable above 90% prediction accuracy. It can conclude that the features of walking speed, downlink and remaining time are utilised to distinguish the difference in distance, and the direction feature can help the system further distinguishes the difference when PCs are in the same group. Combining these features can significantly improve the prediction accuracy.

4.4.3 Prediction Accuracy with Different Randomness

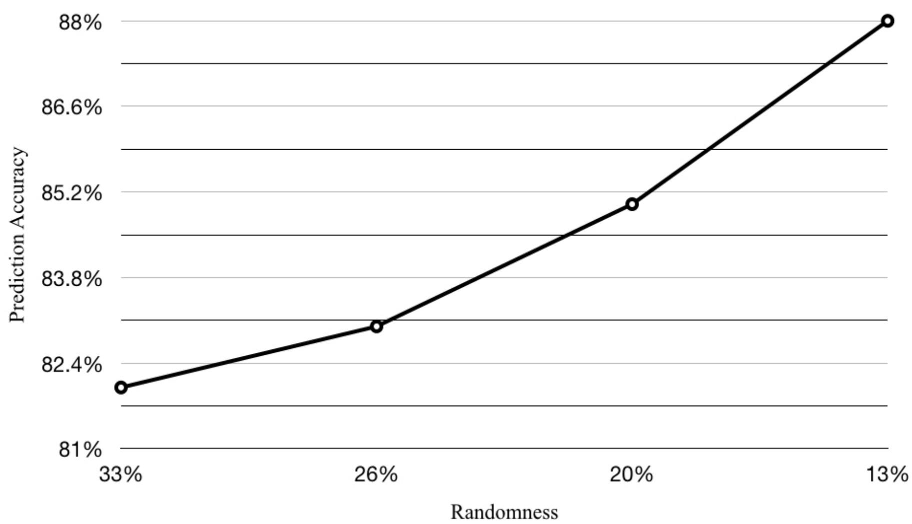


Fig. 4.13 Prediction accuracy with uncorrelated feature.

The prediction can be made based on the assumption that data should have a certain pattern. To destroy the pattern, the number of random moving subscriber data will be increased in this section, and the prediction accuracy should be higher with less randomness. The random user will not contain a certain moving pattern, it means that the user will walk in high speed, and it requires a significant downlink service. Moreover, the destination of user will not related to the user's remaining time. When this kind of random user data is used to generated the pattern, it will not match the original pattern, and the combined pattern will be different than before. When this kind of pattern is used to predict the user destination, the impact will be severe. In the simulation, the randomness is set 33% at the simulation beginning, and 20% random data will be decreased in each simulation, and the prediction accuracy is shown in the Figure 4.13. According to the figure, the system can provide high prediction accuracy with less random data, and the prediction accuracy can achieve 88% when the data set contains 13% random data. The prediction accuracy cannot achieve 100% even no random data within the data set because some PCs still have similar features value.

4.4.4 Predicted Traffic Distribution

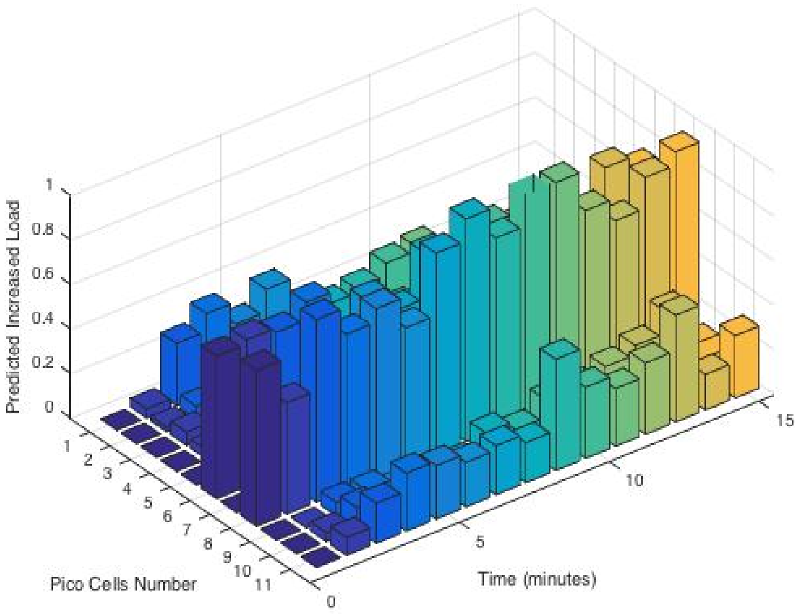


Fig. 4.14 Predicted incoming traffic of pico cells.

The subscriber's walking speed has already collected by the system. Therefore, the system can inform the PC about the incoming traffic with specific timing given the distance between two PCs is known. Figures 4.14 shows the predicted traffic distribution via NBC. The figure shows the traffic demand in some PCs is higher than other PCs because these PCs are located in the area that it will receive traffic from two starting PCs. Some of the PCs do not have high traffic demand because of location is in the scenario edge, and they do not have the common route from two starting PCs. Moreover, the figure shows that the traffic is heavy in some specific time slot because subscribers have different walking speed, and they might arrive at some PCs in the same time slot. Due to the prediction error, it will cause the difference between predicted traffic distribution and actual traffic distribution. Based on the simulation, the total number of resources distribution error is 380. Since the system will give PCs about the wrong incoming traffic information, the power modification cannot perfectly achieve the objective.

4.5 Transmission Power Modification via Traffic Prediction

4.5.1 Transmission Power Modification Process

Traffic prediction and power modification will be combined to reduce network energy consumption, the process is shown in the Figure 4.15. According to the traffic prediction result in the previous section, the system will inform the corresponding PC to increase the transmission power to prepare the incoming traffic. The collection, analysis and prediction is the same with Naive Bayesian classifier process. The power modification will focus on the power modification based on the prediction result from NBC, the system will inform the BS the incoming traffic with a specific timing. If the current power level is not enough, the PC will increase the transmission power to prepare the incoming traffic, otherwise the power will not be modified.

4.5.2 Load Adaptive Power Model

Operating all PCs at full power during light-traffic demand period will waste energy. The load adaptive power model is proposed to modify the transmission power proportional to the load of PC.

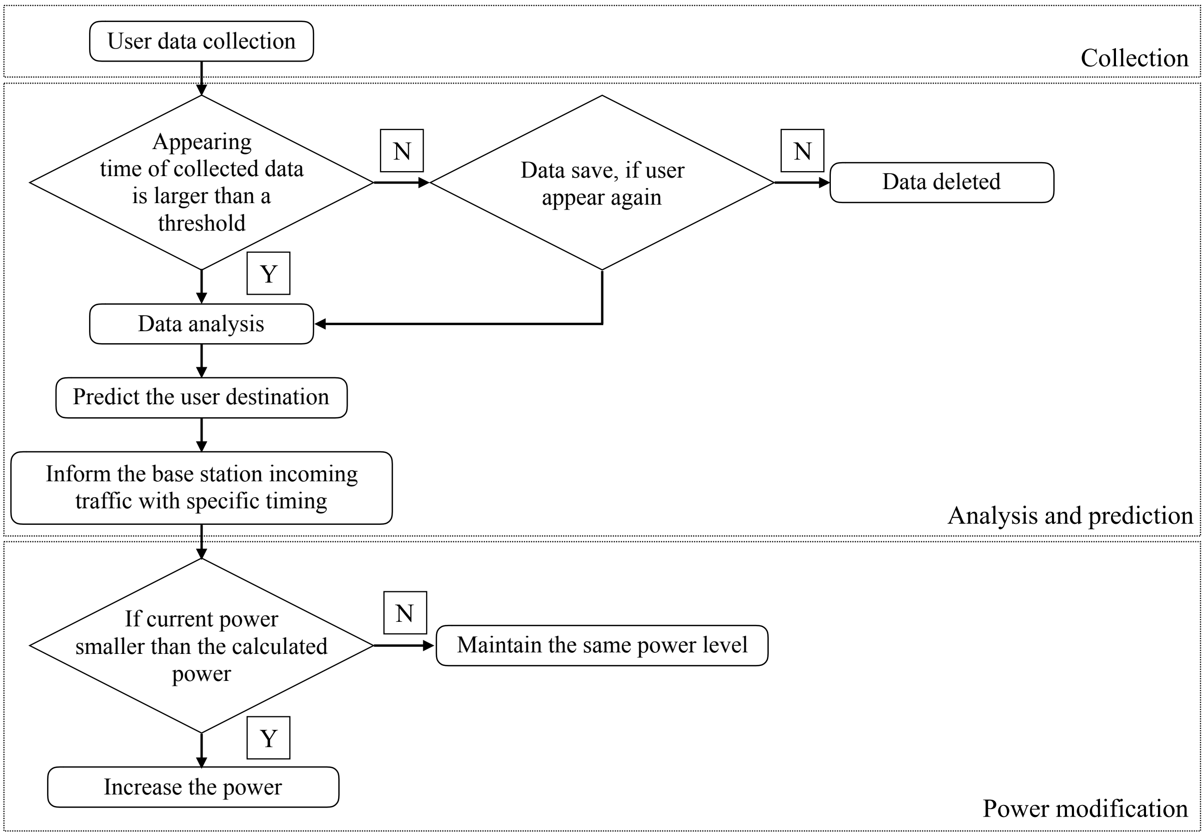


Fig. 4.15 Transmission power modification process via traffic prediction.

Load Model Let's define the load model for the further utilisation in load adaptive power model, the required downlink of subscriber is defined as D_i, $SINR_{O_i}$ represents the SINR of subscriber. Therefore, by following the Shannon channel capacity equation, the required bandwidth B_{O_i} can be obtained, which is shown in Equation 4.17.

$$B_{O_i} = D_i / \log_2(1 + SINR_{O_i}) \tag{4.17}$$

Let's define the load of PC as the total allocated resources of subscribers divided by the total resources number of PC, which is shown in Equation 4.18.

$$L_p = \frac{\sum_1^i X_i B_{O_i}}{R_{tp}} \tag{4.18}$$

R_{tp} represents the total resources of PC. X_i will equal to 1 if subscriber O_i is connected to PC p, 0 otherwise.

Load Adaptive Power Model The power of PC will be modified based on the prediction result from NBc. For energy saving purpose, the transmission power of PC will be set proportional to the load of PC. Let's define the incoming traffic of PC p as L'_{pt} in specific timing t. Therefore, the incoming traffic of PC in t time slot can be represented as $L_{pt} = L_{p(t-1)} + L'_{pt}$. Let's define the transmission power as P_t. Follow the load adaptive model, the load adaptive power of PC can be obtained, which is shown in Equation 4.19 [114].

$$P_p = L_p \mu P_t + (1-\mu) P_t \tag{4.19}$$

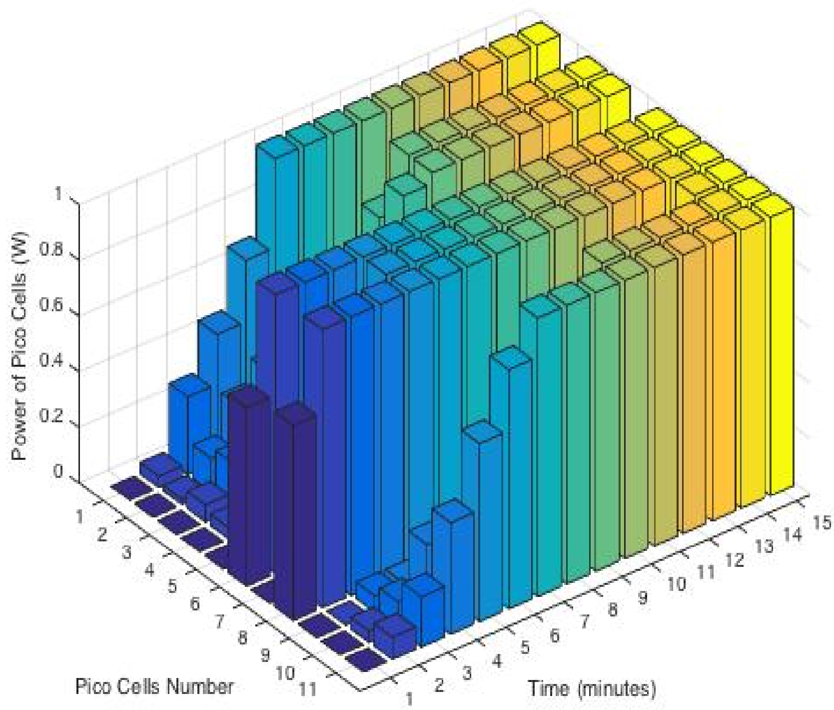


Fig. 4.16 Power of pico cells.

where μ represents the load adaptive factor. According to the equation, it can be seen that the higher value of the μ, the load of PC can cause deeper influence on power. PC will consume a fixed energy if $\mu = 0$, and the transmission power will be completely influenced by the load when $\mu = 1$. Moreover, the energy consumption can be significantly reduced because some PCs can be completely switched off if the load is 0. Therefore, the system only needs to increase the transmission power when traffic demand is growing. The advantage of adding this load adaptive factor is that the system or the technician can modify the factor to increase or reduce the correlation between transmission power and the load.

According to the prediction result, Figure 4.16 shows the transmission power of PC when the load adaptive factor is equal to 1 in the simulation. The value of time axis begins to increase after the simulation starts, and the value is match with the remaining time in the reverse way. It means that the 15 mins remaining time is equal to 1 of the time axis in Figure 4.16. The power of pico cells axis shows that value of the pico cells' power when the time goes in the simulation. The trend of the increased power will proportional to the increased cell load to serve increasing number of user.

4.5.3 Operated Transmission Power and Load

Load of Pico Cells The Figure 4.17 shows the load of PC during the simulation, and the number 12 represents the load of MC. The load of each PC continue growing when increasing number of subscribers appear within the network. According to the Figure 4.17, it can be seen that most of the PCs are approaching 100% load at the end of the simulation. Moreover, the load of MC is increased fast at the beginning of simulation because most of the PCs are operated in low power level. When subscribers are moving to their destination, subscribers will be served by the MC rather than the closest PC. The load of MC will be reduced after the power of PC is back to regular level, the coverage of PC is expanded to absorb the MC's traffic.

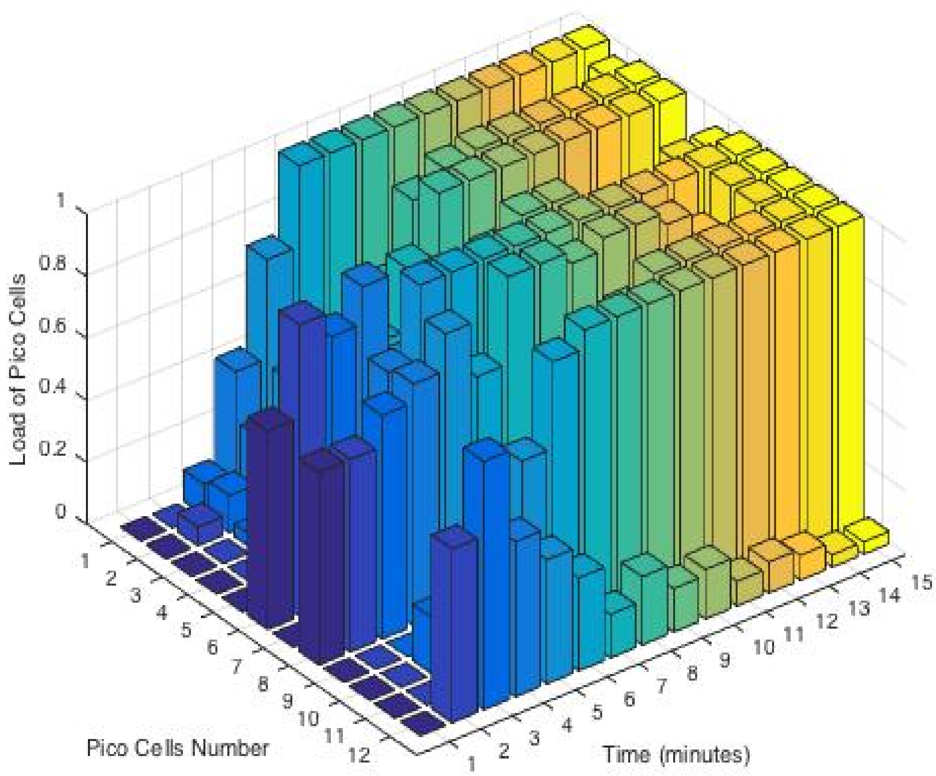


Fig. 4.17 Load status of pico cells.

Operated Transmission Power According to the traffic prediction result, the predicted traffic distribution in each time slot can be obtained. The system will apply the load adaptive power model to modify the PC's transmission power. The modified transmission power of each PC is shown in Figures 4.18-4.23, from time slot 1 to 5. The reason of focusing on PCs 6 and 8 is mainly because high-traffic area only occurs in the coverage of PCs 6 and 8 in the beginning of the simulation, and the subscribers will first appear in these PCs. It means the load of PCs 6 and 8 will continues grow up. When simulation time goes, user will move and end in the remaining PCs, causing the increased load of the remaining PCs At the beginning of the simulation, the transmission power of two PCs is set at a medium level. In the second time slot, the subscribers will begin to move to the nearby PCs. Therefore, some of the PCs near PCs 6 and 8 begin to increase the transmission power to serve the incoming subscribers.

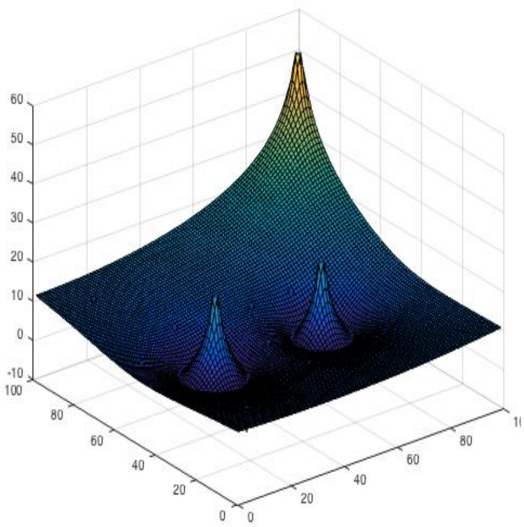


Fig. 4.18 Transmission power of pico cells in first time slot.

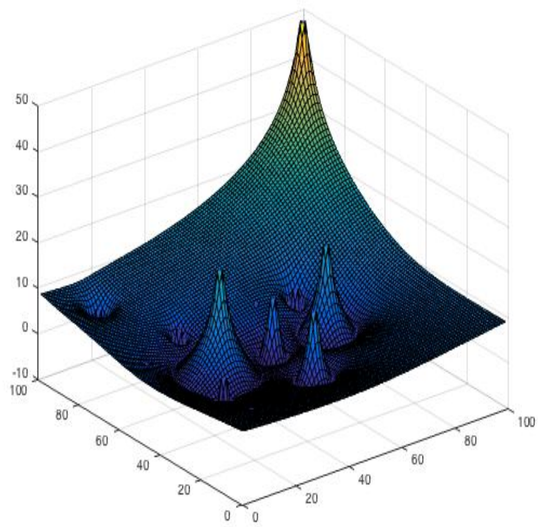


Fig. 4.19 Transmission power of pico cells in second time slot.

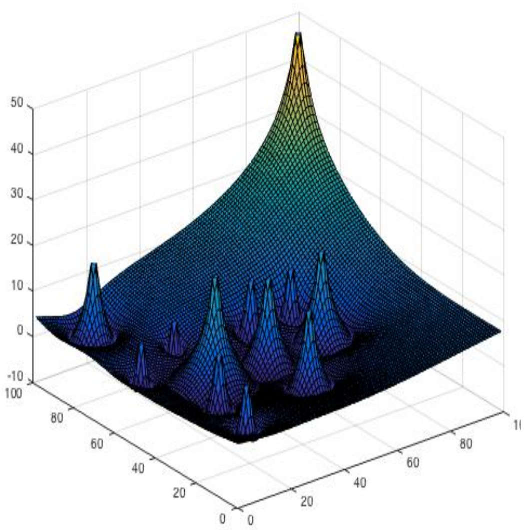


Fig. 4.20 Transmission power of pico cells in third time slot.

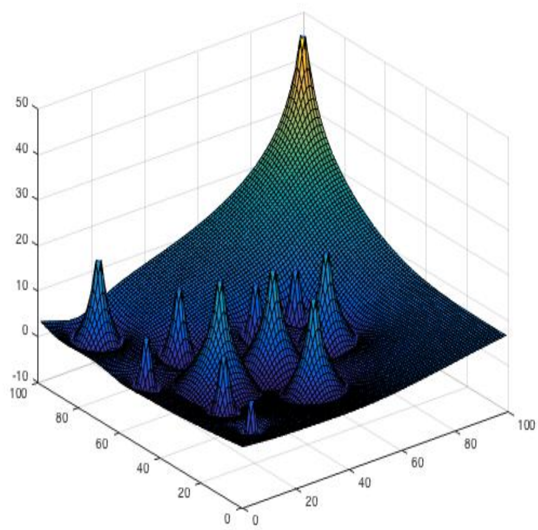


Fig. 4.21 Transmission power of pico cells in fourth time slot.

Moreover, some of the high walking speed subscribers might move to a remote PCs, which is PC 1 in the corner, the system will inform the corresponding PCs to slightly increase the transmission power. In the following time slots, because subscribers continue to appear in

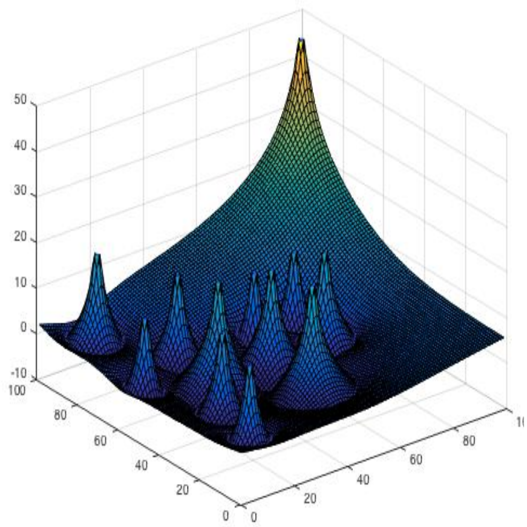


Fig. 4.22 Transmission power of pico cells in fifth time slot.

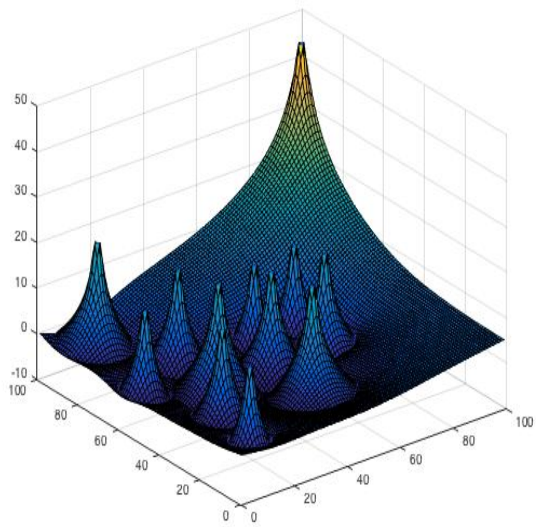


Fig. 4.23 Transmission power of pico cells in sixth time slot.

the starting PCs, the traffic demand in PCs 6 and 8 are still in the medium or high level in the simulation, and the transmission power is modified proportional to the load of their load. It is worth to notice that the traffic demand in the overlapping area is heavy during the simulation because traffic will come from PCs 6 and 8, the PCs within this area will be operated at the high power. However, some of the remote PCs might not operated in full power.

4.5.4 Energy Consumption and Energy Efficiency

The Figure 4.24 and 4.25 show the total energy consumption of PCs and the EE comparison after the strategy is applied in the network.

Energy Consumption The load adaptive power model is applied to change the transmission power of PC, the total energy consumption of PCs will be different with different value of load adaptive factor. The lines in Figure 4.24 represent the total energy consumption of PCs of the original and the factor equal to 0.5 and 1, which are represented by blue, orange and yellow lines. From the figure, it can obtain the energy consumption difference at any

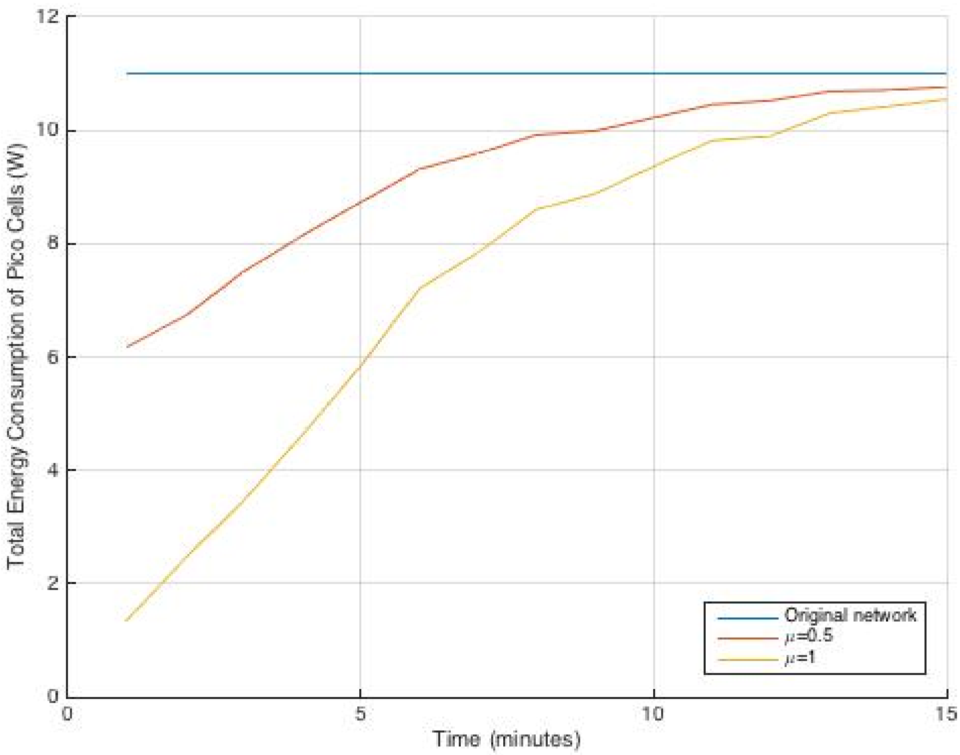


Fig. 4.24 Total energy consumption of pico cells.

time with any load adaptive factor. Moreover, the figure can also tell the energy consumption difference of the entire simulation by calculating the area difference between two lines.

Meanwhile, the figure also tells some other information. For example, orange and yellow lines have lower energy consumption than blue line, more energy can be saved when the load adaptive factor is increased, especially in light-traffic period, which is shown in the figure from 1 to 5. In the figure, the energy consumption of original network in time slot 1 is 11, and the energy consumption of red line and yellow line in time slot 1 is 1.34 and 6.17. Therefore, it can obtain the energy consumption difference are 9.66 and 4.83. However, the energy reduction will become smaller due to the traffic demand is increased. The PCs shall be operated in normal transmission power to serve the increasing traffic. At the end of the simulation, the energy consumption of red line and yellow line are 10.77 and 10.54. The energy consumption difference with the original network is 0.23 and 0.46.

Therefore, the energy consumption difference is limited when the simulation reach the end. It can be seen that the energy reduction is significantly during the traffic demand is light. The load adaptive power model cannot reduce too much energy consumption when traffic demand is heavy. However, it can save 33% of the energy in the entire simulation length.

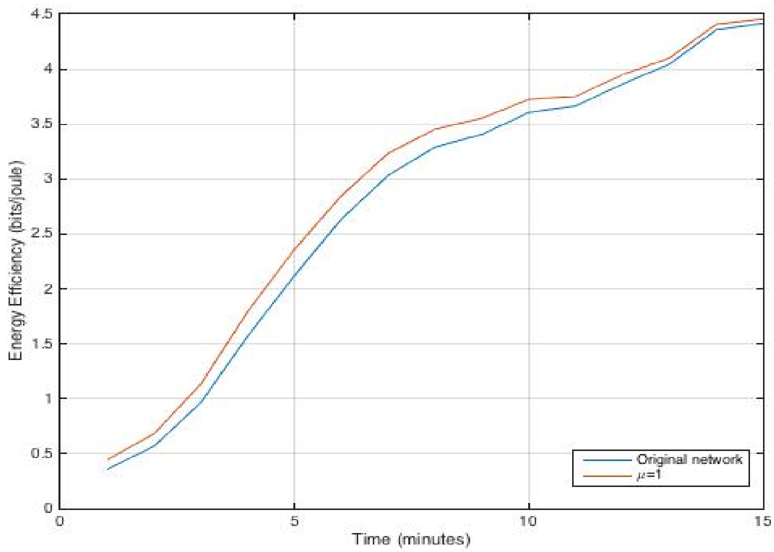


Fig. 4.25 Energy efficiency.

Energy Efficiency Follow the EE model in chapter 3, section 3.2.7. It defines the EE as the total network throughput divided by the total network energy consumption, which is calculated as bits per joule. The EE comparison is made in Figure 4.25 to show the EE difference between original network and load adaptive factor is applied in the simulation. The blue line represents the EE in the original when all PCs are operated in full transmission power, and load adaptive factor is set as 1, which is represented by the red line. According to Figure 4.25, it can be seen that the load adaptive factor can significantly influence the energy consumption. In the original network, the EE in first time slot is 0.35, and 0.44 with factor value 1, the improvement is significant. Because MC is operated to provide coverage, and the power of PC can be set as low, the subscribers who move outside the PCs 6 and 8 coverage can be served by the MC. In the following time slots, the transmission power of some PCs should be gradually increased to let more PCs to serve the increasing number of subscribers, and the EE difference will be only 0.04 at the end of the simulation. It can be seen that the load adaptive power model cannot work efficiently in high-traffic demand scenario. However, the EE improvement can achieve 4.7% during the entire simulation, and most of the improvement is achieved in light and mid traffic period, which is time slot 1 to 12.

To conclude, the simulation shows that the load adaptive power model can sufficiently work in low and mid traffic demand scenario, the EE improvement is significant. However, when the traffic demand is increased, all PCs need to increase transmission to serve subscriber, and the effort of load adaptive power model is limited.

4.5.5 Energy Efficiency in Different Scenarios

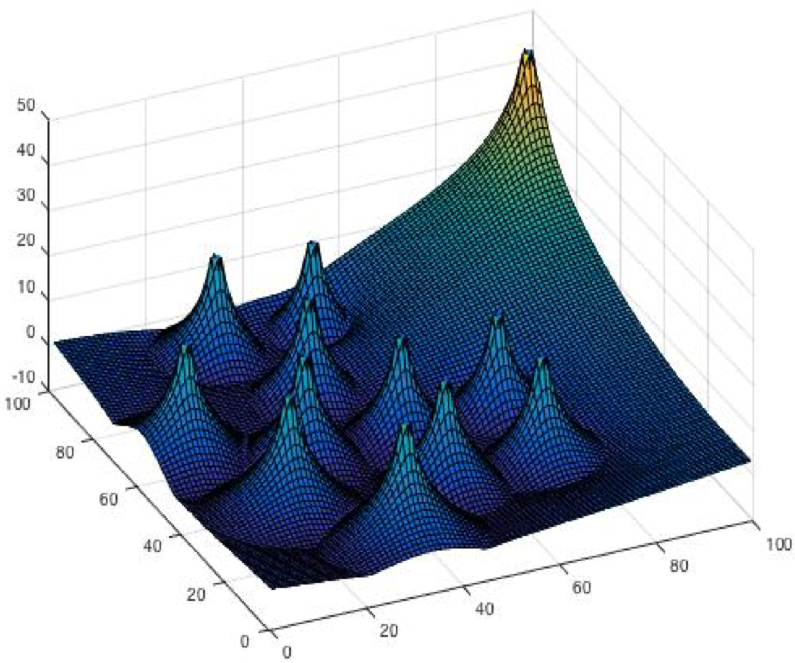


Fig. 4.26 Scenario two.

Two more scenarios are put into test to estimate the EE of the proposed method. Except the starting PCs are remain the same with the previous scenario, the paths and the destination of the subscribers are reset. The next two scenarios will follow the same method to predict the subscriber's destination and modify the power of PCs.

Scenario Two The scenario two is shown in Figure 4.26, and the paths between PCs and subscribers are changed, the detail of the paths is shown in Table 4.5.

Table 4.5 Pass pico cells between starting pico cell and ending pico cell in scenario two.

Pass Pico Cells		
Starting pico cell	Ending pico cell	Pass pico cell number
6	1	6-4-1/6-2-1
6	2	6-2
6	3	6-4-3
6	4	6-4
6	5	6-5
6	9	6-10-11-9/6-5-9
6	10	6-10
6	11	6-10-11
8	3	8-10-4-3/8-7-3
8	4	8-10-4
8	9	8-7-11-9/8-10-11-9
8	10	8-10
8	11	8-10-11/8-7-11

The EE of scenario two is shown in Figure 4.27. The blue colour line represents the EE of the original network, and the red colour line represents the EE of the network when the transmission power is proportional to the predicted load of the PCs. Compared with the original network, the EE improvement of the total network is 4.4% with 33% energy saving.

Scenario Three The scenario three is shown in Figure 4.28 with the new paths between PCs and sucscribers, the detail of the paths is shown in Table 4.6.

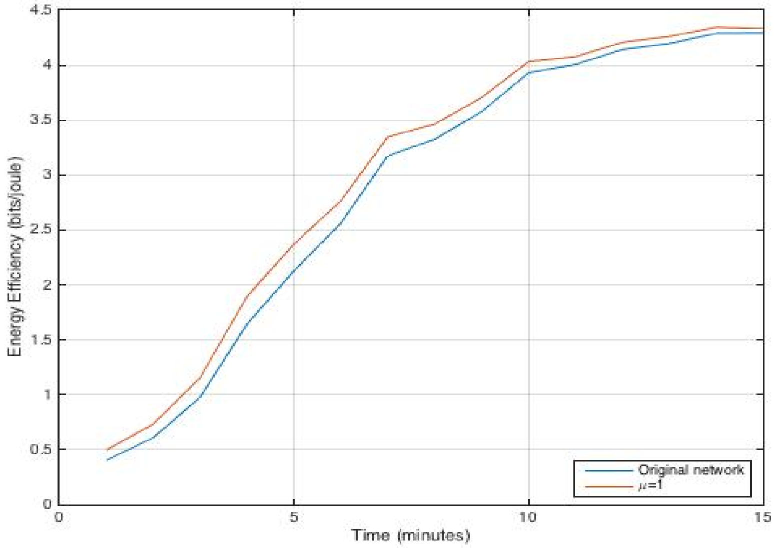


Fig. 4.27 Energy efficiency of scenario two.

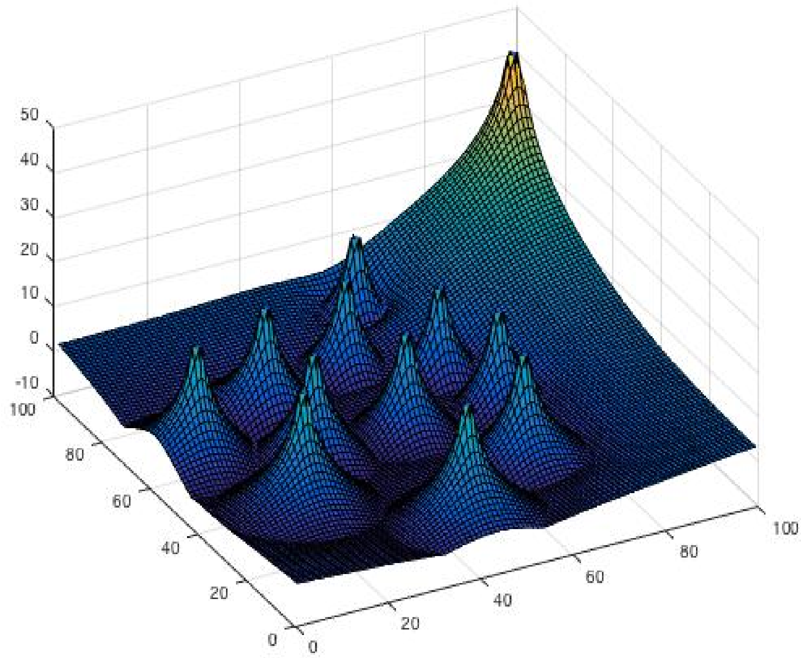


Fig. 4.28 Scenario three.

The EE of scenario three is shown in Figure 4.29. The blue colour line represents the EE of the original network, and the red colour line represents the EE of the network after the

Table 4.6 Pass pico cells between starting pico cell and ending pico cell in scenario three.

Pass Pico Cells		
Starting pico cell	Ending pico cell	Pass pico cell number
6	2	6-2
6	3	6-2-3/6-7-3
6	5	6-5
6	7	6-7
6	9	6-7-9/6-10-9
6	10	6-10
8	1	8-4-1
8	3	8-4-3/8-7-3
8	4	8-4
8	7	8-7
8	9	8-7-9/8-11-9
8	11	8-11

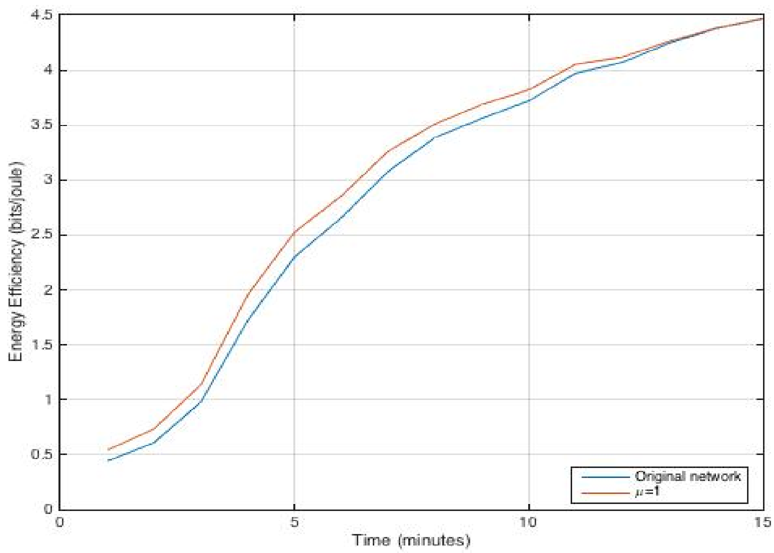


Fig. 4.29 Energy efficiency of scenario three.

transmission power is modified according to the predicted load of the PCs. It can be seen that compared with the original network, the EE improvement of the total network is 3.9%. Meanwhile, it can save 30% of the total network energy consumption.

In three different scenarios, the EE improvement is 4.7%, 4.4% and 3.9%, the average improvement is 4.3%. The reason of causing the EE improvement difference depends on the location of the PCs and the paths. The number of remote PC will cause some impacts on the total energy consumption is mainly because the closed PCs shall be operated to serve the incoming subscribers, these subscribers will go to both remote and closed PCs. In three scenarios, there are only two remote PCs in scenario three, and three remote PCs in scenarios one and two. Therefore, more closed PCs will force the network to increase the total network power rapidly due to the closed PCs will be the first serving BS for short or long distance subscribers. Moreover, another minor impact comes from the paths because the chance of reaching every PC is the same, it means if more than one PCs can take part of the traffic from the nearby PCs, the operated power of PC can be reduced. However, the EE improvement is still significant, and the simulation results prove that the proposed method can sufficiently reduce the energy consumption of the network, improving the EE.

4.6 Conclusions

In this chapter, it proposes the NBC to collect and utilise subscriber's data to predict the traffic distribution in CN, the features of walking speed, downlink, remaining time and direction are collected when subscriber connect to the network. After some subscribers' data is filtered by the system, the remaining subscribers' data will be analysed and utilised for the traffic prediction, and it can determine a strong correlation between subscriber features and PC. The walking speed, downlink and remaining time can help the system to distinguish the distance difference between starting location and ending location. However, these features can not help the system to distinguish the difference when some PCs are classified in the same group. Therefore, direction feature is added to improve the prediction accuracy. After the direction is added into prediction, the system can distinguish the difference even some PCs are in the same group. Based on the simulation, it can conclude that if the prediction is made by the single feature, the prediction accuracy is unsatisfied. However, if it put all features into traffic prediction, significant enhancement in prediction accuracy can be obtained.

After the system obtains the traffic prediction result via NBC, the system can foresee the traffic growing in each PC. If some PCs only need to serve a small number of subscribers, these PCs can be operated in low power. Therefore, load adaptive power model is proposed to modify the transmission power of PC. The power of PCs will be increased proportional to the predicted load of PCs to match the growing traffic demand. During light-traffic period, some of the PCs can be switched off for reducing energy consumption, the subscribers can be served by the MC rather than the PC. When the network traffic demand is continues growing, the system will inform the corresponding PC to increase transmission power. According to the simulation results, it shows that it can significantly reduce the energy consumption of network, and it also shows that the EE can be improved when the traffic demand is light, the load adaptive power model cannot perform well in high-load scenario.

Chapter 5

Spectrum Efficiency Improvement via Joint Transmission and Cell Range Expansion

5.1 Introduction

The advanced wireless technology causes the fast-growing traffic demand in the CN. In order to match the demand, different type of low power BS is deployed within HeNet to increase the network capacity and extra access points. When subscriber appears in cell edge, weak received signal strength and high interference from other BSs degrade the user experience. Moreover, The interference in HeNet will become more severe because multiple BSs from different tier might generate interference to the subscriber in downlink. Therefore, cell edge performance will become a challenge to network operators.

To tackle worst downlink performance in cell edge, JT is proposed to mitigate the interference from other BSs. This technique will inform multiple BSs to transmit signal to one subscriber to improve the downlink performance, typically the cell edge performance. Subscriber in cell edge will simultaneously receive signals from two or even more BSs to enhance *SINR*. Therefore, JT requires same amount of resources from serving BS and cooperative BS. However, the traffic is not uniformly distributed within the network, resources might have already assigned to the cell central subscriber, cooperative BS might not have sufficient resources to execute JT. If the system force cooperative BS to allocate resources for JT, central subscribers cannot receive sufficient resources. Not only the subscriber cannot get satisfied service from the network but also the overall network performance degradation. Moreover, if subscribers chooses a remote BS as cooperative BS, cooperative BS can only provide a weak signal to subscriber, the improvement is limited. Therefore, the problem of insufficient resources in cooperative BS should be tackled in JT.

The uneven traffic distribution might cause the problem of traffic congestion in cooperative BS, executing JT might reduce the overall network performance. The unbalanced traffic demand could reduce the resources utilisation and user experience, balancing the traffic is considered as one of the key features in SoN. Meanwhile, some BSs only serve for a small number of subscribers, when high-load BS appears in the network, many options can be chosen, offloading the traffic to neighbour BSs or low-tier BSs. Due to the difference in transmission power, subscriber might have higher chance to connect to the MC, and the traffic cannot sufficiently offload from MC. CRE is proposed to offload the traffic from MC to low-tier BSs [115] - [118], the bias is added into user association to change the serving BS of subscriber without increasing the transmission power. In [119], the research shows that the optimal performance can be achieved with carefully chosen bias in CRE. The coverage can be changed when the power of PC is increased or large bias is added into cell selection scheme [120]. Therefore, CRE can be utilised to tackle the traffic congestion in cooperative BS when resources are insufficient for JT, the cell edge performance and resources utilisation can be improved without degrading overall network performance, and subscriber does not need to choose remote BS as cooperative BS. The system chooses MC as cooperative BS, after traffic sufficiently is offloaded to low-tier BS, JT can be frequently applied to improve the cell-edge user downlink performance. If PC is chosen to participate JT, bias can be modified to reduced the offloading traffic ratio from MC. Therefore, the system should modify the bias based on the BS tier of cooperative BS. However, the load balance problem will become a NP-hard problem with large number BSs and subscribers, the flow water algorithm is applied in MC to solve load balancing problem, the algorithm will offload the traffic to the lowest-load BS, CRE bias will be estimated after the algorithm has chooses the target BS. The algorithm will be stopped until the load difference reaches minimum value. The simulation will show the gain from JT after the traffic congestion is tackled.

The remaining parts are organised as the following structure. Section 5.2 will describers the system model. Section 5.3 will introduce the JT, including the the problem of insufficient resources in cooperative BS. CRE will be introduced in section 5.4. Section 5.5 and section 5.6 will show the offload traffic via CRE, and the gain from JT. Section 5.7 will give the conclusion of this chapter.

5.2 System Model

5.2.1 Notation

Let's consider a HeNet is deployed with multiple MCs and PCs, Let's denote a set of MCs and PCs, where $\{1,2,3,...m\} \in M$ and $\{1,2,3,...p\} \in P$. Meanwhile, let 's denote the interference from MCs and PCs, where $\{1,2,3,...m\} \in M'$ and $\{1,2,3,...p\} \in P'$. The subscriber group is defined as $\{1,2,3,...i\} \in I$. In JT, let's consider that the cooperative BSs from MCs $\{1,2,3,...m\} \in M_{co}$ and PCs $\{1,2,3,...p\} \in P_{co}$.

5.2.2 Signal to Interference Plus Noise Ratio

Let's consider the subscribers i is connected to BS, which is determined by the strongest received signal strength [121]. The received signal strength of subscriber is estimated by path loss model, and the model is described below.

Free Space Path Loss Model Free space path loss model is utilised in this chapter to estimate the transmission loss in free space, the equation of free space path loss can be represented by the following equation in terms of dB

$$\begin{aligned} FSPL &= 10log_{10}((\frac{4\pi df}{c})^2) \\ &= 20log_{10}(\frac{4\pi df}{c}) \\ &= 20log_{10}d + 20log_{10}f + 20log_{10}(\frac{4\pi}{c}) \\ &= 20log_{10}d + 20log_{10}f - 147.55 \end{aligned} \tag{5.1}$$

where d is measured in units of metres and f is measured in units of hertz. c represents the speed of light. Therefore, received signal strength in subscribers side can be obtained given the transmission power is known.

SINR The *SINR* equation is shown in Equation 5.2

$$SINR = \frac{S}{I+N} \tag{5.2}$$

where S represents the received signal strength at subscribers location, which is estimated by the free space path loss model. I represents the interference from the remaining BSs, and N represents the background noise. Therefore, according to the Equation 5.2, *SINR* of subscriber i which is connected to serving MC m or PC p can be represented by the following equation.

$$SINR_{im} = \frac{P_{im}G_{im}}{\sum_{m=1}^{M'} P_{im}G_{ip} + \sum_{p=1}^{P'} P_{ip}G_{im} + \sigma^2} \tag{5.3}$$

$$SINR_{ip} = \frac{P_{ip}G_{ip}}{\sum_{m=1}^{M'} P_{im}G_{im} + \sum_{p=1}^{P'} P_{ip}G_{ip} + \sigma^2} \tag{5.4}$$

where the transmission power of MC and PC are denoted as P_{im} and P_{ip}. G_{im} and G_{ip} represent the channel gain when the subscriber connects to the corresponding BS, which includes path loss only. σ^2 denotes the noise power. Therefore, it can obtain the subscribers *SINR* in any location within CN, and it will be used to estimate the downlink performance and calculate the throughput.

SINR in Joint Transmission Let's assumed that up to n BSs will transmit signal to one subscriber simultaneously. According to Equation 5.2, S represents the received signal strength. Therefore, the subscriber will receive signal from MC m, and more than one BS's (cooperative BS) signal in JT, when the system chooses MC as cooperative BS, the $SINR_{imm}$ can be represented as

$$SINR_{imm} = \frac{P_{im}G_{im} + \sum_{m=1}^{M_{co}} P_{im}G_{im}}{\sum_{m=1}^{M'} P_{im}G_{im} + \sum_{p=1}^{P'} P_{ip}G_{ip} + \sigma^2} \tag{5.5}$$

The *SINR* expression can be obtained when JT is executed between MC and PC, which is shown in Equation 5.6. Moreover, after the role of serving BS and cooperative BS is exchanged, serving BS m and cooperative BS p or the reverse way. The following equations can be obtained, which is shown in Equation 5.7.

$$SINR_{imp} = \frac{P_{im}G_{im} + \sum_{p=1}^{P_{co}} P_{ip}G_{ip}}{\sum_{m=1}^{M'} P_{im}G_{im} + \sum_{p=1}^{P'} P_{ip}G_{ip} + \sigma^2} \tag{5.6}$$

$$SINR_{ipm} = \frac{P_{ip}G_{ip} + \sum_{m=1}^{M_{co}} P_{im}G_{im}}{\sum_{m=1}^{M'} P_{im}G_{im} + \sum_{p=1}^{P'} P_{ip}G_{ip} + \sigma^2} \tag{5.7}$$

5.2.3 Throughput

According to the Shannon channel capacity equation [122] [123], which is shown in the following equation, the subscriber's throughput can be obtained.

$$C = B \times \log_2(1 + SINR) \tag{5.8}$$

where B represents the allocated bandwidth of subscriber. Let's denote the throughput of subscribers i as T_i, the throughput can be represented by the following equation.

$$T_i = R_{im} log_2(1 + SINR_{im}) \tag{5.9}$$

where R_{im} represents the allocated resources. Therefore, the throughput of JT can be represented by Equation 5.10 - 5.12 with respective serving BS and cooperative BS.

$$T_{imm} = R_{im} log_2(1 + \frac{P_{im}G_{im} + \sum_{m=1}^{M_{co}} P_{im}G_{im}}{\sum_{m=1}^{M'} P_{im}G_{im} + \sum_{p=1}^{P'} P_{ip}G_{ip} + \sigma^2}) \tag{5.10}$$

$$T_{imp} = R_{im} log_2(1 + \frac{P_{im}G_{im} + \sum_{p=1}^{P_{co}} P_{ip}G_{ip}}{\sum_{m=1}^{M'} P_{im}G_{im} + \sum_{p=1}^{P'} P_{ip}G_{ip} + \sigma^2}) \tag{5.11}$$

$$T_{ipm} = R_{ip} log_2(1 + \frac{P_{ip}G_{ip} + \sum_{m=1}^{M_{co}} P_{im}G_{im}}{\sum_{m=1}^{M'} P_{im}G_{im} + \sum_{p=1}^{P'} P_{ip}G_{ip} + \sigma^2}) \tag{5.12}$$

5.2.4 Load Model

The load model is followed the model in Chapter 3 [89]. Let's define the load of BS as occupied resources divided by total resources of BS. Let's set the load of MC and PC as L_m and L_p, the load can be represented by Equation 5.13 and 5.14.

$$L_m = \frac{\sum_{i=1}^{I} R_{im}X_{im}}{R_{tm}} \tag{5.13}$$

$$L_p = \frac{\sum_{i=1}^{I} R_{ip}X_{ip}}{R_{tp}} \tag{5.14}$$

where R_{im} and R_{ip} represent the allocated resources in respective BS. Total resource block of MC and PC are denoted as R_{tm} and R_{tp}. X_{im} and X_{ip} equal to 1 when the subscriber i is connected to the respective BS, otherwise 0.

5.3 Joint Transmission with Different Cooperative Base Station

5.3.1 Scenario

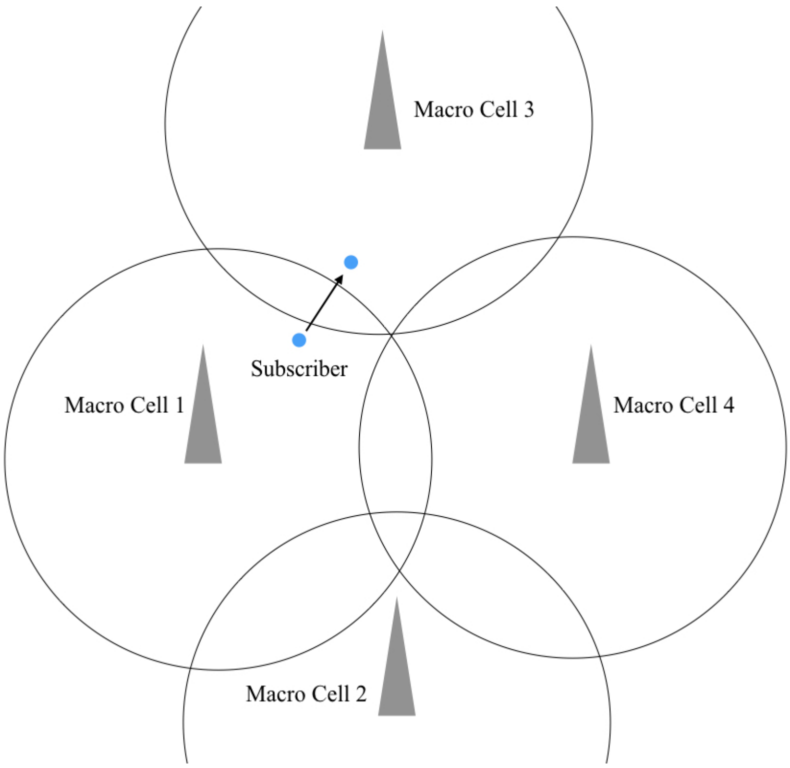


Fig. 5.1 Scenario.

Let's consider the scenario that four MCs are deployed within the network, and the subscribers are uneven distributed within the scenario. The scenario is shown in Figure 5.1. One subscribers appear in MC 1 edge, which is shown in Figure 5.1 as blue dot. The subscriber begins to move to the MC 3 centre, the direction is shown in Figure 5.1 as arrow line. 8 movements will be simulated in the simulation, the subscriber will move one point in one time slot. Moreover, for the analysis simplicity, the remaining will be considered as stable during the simulation. During the movement, the received signal strength from the serving MC is getting worse.It is assumed that the downlink performance can not match the subscriber's demand, JT need to be executed to improve the downlink performance. The remaining 3 MCs will become the potential cooperative BSs, but the load of each MC is different, which is shown in Table 5.1. Two MCs are suffer from traffic congestion, it means that these MCs cannot provide sufficient resources for executing JT.

Table 5.1 Load of macro cells.

Macro cells load	
Macro cells Number	Load
1	66%
2	95%
3	100%
4	34%

5.3.2 Simulation Result

The following simulation results will show the JT gain and throughput change in respective cooperative BSs with three different scenario. The JT gain and throughput lost when MC 2 is selected as cooperative BS are shown in Figure 5.2 and 5.3. JT gain and lost when MC 3 is selected as cooperative BS are shown in Figure 5.4 and 5.5, JT gain when MC 4 is selected as cooperative BS are shown in Figure 5.6.

Macro Cell 2 According to the location of MC 2 in Figure 5.1, it is the most remote BS to the moving subscriber, and Figure 5.2 shows the gain from JT if the system chooses MC 2 as cooperative BS, the moving subscriber can obtain a certain throughput gain from JT, but the gain has a decrease trend. Moreover, the subscriber is moving to the opposite direction of MC 2, and the received signal strength will getting weaker due to the increasing distance between BS and subscriber, and it causes the decreased JT gain given the subscriber's allocated resources is unchanged.

Meanwhile, 95% load in MC 2 indicates that the resources shortage issue occurs in cooperative BS. If the system force MC 2 to participate JT, the allocated resources for central subscriber will reassign to JT subscriber, and it will cause the throughput loss due to the spectrum efficiency better between the central subscriber and cell-edge subscriber, and the throughput loss in MC 2 is shown in Figure 5.3. The increased throughput loss is caused by the higher path loss when subscriber is moving out the coverage of MC 2.

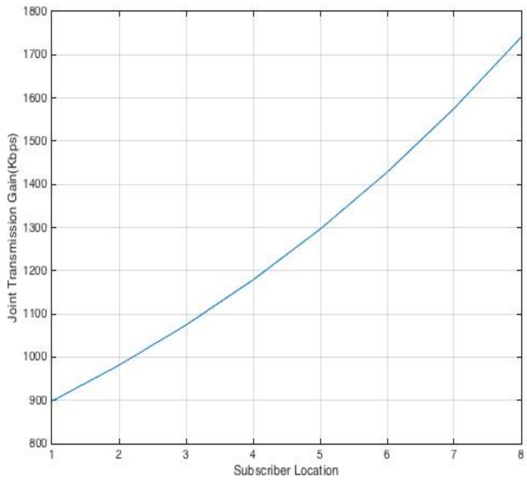


Fig. 5.2 Joint transmission gain from macro cell 3.

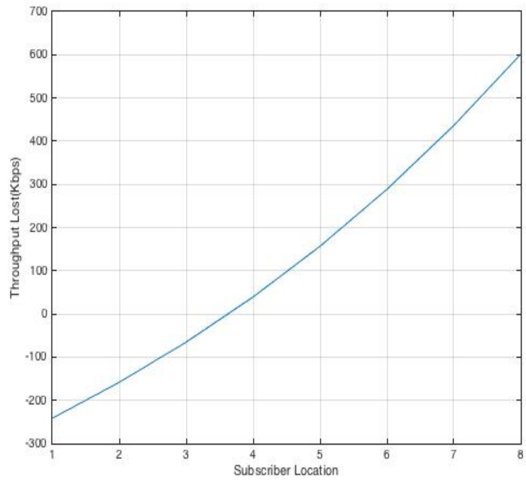


Fig. 5.3 Throughput lost with cooperative macro cell 3.

Macro Cell 3 If the cooperative BS is determined by the strongest received signal strength from the remaining MCs. According to the scenario, MC 3 will be the first and best option due to the lower path loss than the remaining MCs. Compared with the other MCs, the moving subscriber can obtain the highest gain due to the subscriber can receive optimal received signal strength from cooperative BS and less interference from the other MCs, the gain of JT is shown in Figure 5.4. It can be seen that the gain will be increased when subscriber is moving toward MC 3 centre, obtaining higher spectrum efficiency. The line in Figures 5.4 shows that the trend is completely different with the trends in Figure 5.2 and 5.6 because the subscriber is moving towards MC3.

In the second scenario, the load of MC 3 is 100%, it means that MC 3 will face the same issue as MC 2, the resources is insufficient for executing JT. However, the situation is different with MC 2, and the loss only occur in the first three time slot (1-3) in Figure 5.5. At the beginning, the starting position of moving subscriber is within MC 1 coverage, the received signal strength is weak. Compared with the throughput loss in MC 3, the throughput gain in cooperative BS is insignificant. However, the total network throughput will increase after the time slot 4 because the subscriber is moving to the MC 3 centre, and the received signal strength will be improved due to the path loss is getting small. The increased spectrum efficiency will let the JT gain larger than the throughput loss in cooperative BS. Moreover, the subscriber will keep moving toward the central area of MC 3, the serving BS will change from MC 1 to MC 3.

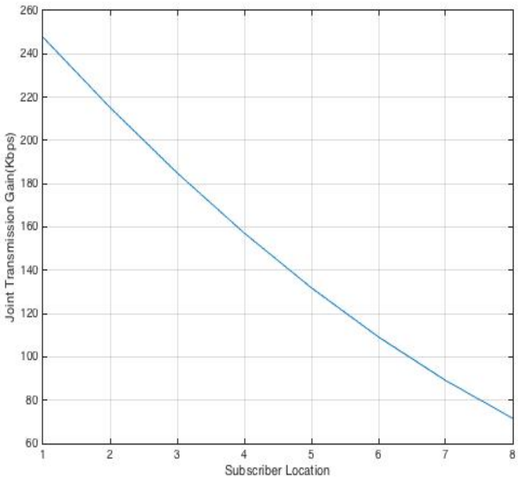


Fig. 5.4 Joint transmission gain from macro cell 2.

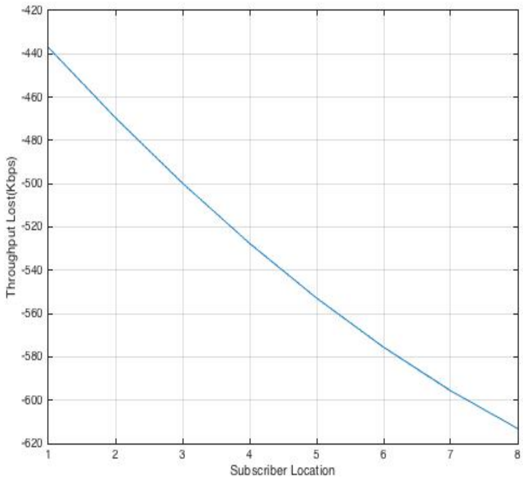


Fig. 5.5 Throughput lost with cooperative macro cell 2.

Macro Cell 4 In the next scenario, the cooperative BS is MC 4, and the resources is sufficient among four MCs, only 34% load. Therefore, MC 4 does not have the problem of traffic congestion, and the sufficient remaining resources can be allocated to the moving subscriber without causing any impact to the serving subscribers. Meanwhile, even cooperative BS

(MC 4) has to face the high path loss, the system can still obtains gain from JT. However, the effort is little, which is clearly shown in Figure 5.6, it is almost the lowest JT gain in three scenarios. Meanwhile, the gain will be decreased due to the subscriber is moving to the MC 3 centre, and not only the increased path loss but also the high interference will be obtained, causing the worse spectrum efficiency.

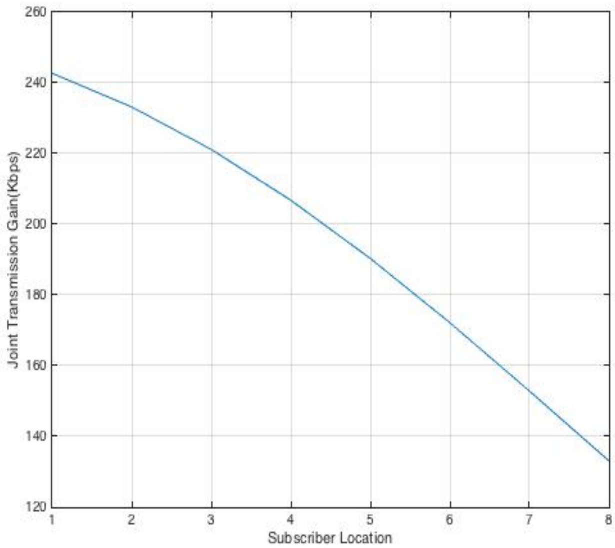


Fig. 5.6 Joint transmission gain from macro cell 4.

It can be seen that if the cooperative BS is determined by the strongest received signal strength, MC 2 and 4 will not be the first and best choice. However, according to the scenario, even MC 3 will become the best option for triggering and participating the JT, the insufficient remaining resources in MC 3 will cause the throughput loss. In this case, resources have to reassign to the subscriber in cell-edge to execute JT, and the worse spectrum efficiency will not bring significant high throughput to the system. It can not fully utilise the existing BS's resources. From the aspect of overall network throughput, even MC 4 will not cause any throughput loss to the system, and subscriber can also obtain gain from JT, but MC 4 provides the lowest spectrum efficiency, considering MC 4 as cooperative BS will not be the best option. Moreover, the problem will become severe if the load of MC 4 is increased. Therefore, it shall find a way to balance the gain of JT and the throughput loss in cooperative

BS. It can be seen that the best solution is to utilise resources of MC 3 to execute JT without any throughput loss in cooperative BS. Therefore, the solutions for reducing the load of MC 3 will be investigated in the following sections.

5.4 Cell Range Expansion for Traffic Offloading

To execute the JT without network throughput degradation, technic shall be applied in cooperative BS to tackle the problem of insufficient remaining resources. CRE is utilised in this chapter the offload the traffic from MC to PCs, the detail will be described in detail in the following section.

5.4.1 Cell Range Expansion

In HeNet, some cell association schemes are based on the Reference Signal Received Power (RSRP), the subscriber will connect to the BS that it can provide the strongest RSRP. Due to the power of MC is much larger than PC, fewer subscribers can be served by the PC. Therefore, the resources of PC cannot be fully utilised [124], and the traffic cannot be sufficiently offloaded from MC to PC. By adding bias in cell selection, the coverage of PC can be expanded without increasing the transmission power [128]. Moreover, bias can be modified to offload more MC's traffic to low-tier BS [120]. Let's denote $RSRP_{BS}$ as the RSRP of BS, the cell association scheme can be expressed as

$$Serving_{BS} = \arg\max_{\{M \cup P\}} \{RSRP_{BS}\} \tag{5.15}$$

The objective of CRE is to offload the traffic from MC to PC in this section, the bias is added to force part of the subscribers connect to the PC [124]. The expression can be represented as

$$Serving_{BS} = \underset{\{M \cup P\}}{\arg\max}\{RSRP_{BS}\} + Bias_{sc} \tag{5.16}$$

$$\begin{cases} Bias_{sc} = 0, & \text{if} \quad \arg\max\{RSRP_{BS}\} \text{in MC tier.} \\ Bias_{sc} \geqslant 0, & \text{if} \quad \arg\max\{RSRP_{BS}\} \text{in PC tier.} \end{cases} \tag{5.17}$$

In Equation 5.16, $Bias_{sc}$ represents the offset for the cell association. $Bias_{sc}$ will equal to 0 if the strongest RSRP is provided by the BS from MC tier, other $Bias_{sc} \geqslant 0$. The larger value of $Bias_{sc}$ will offload more subscribers from MC to PC. The $Bias_{sc}$ calculation equation can represented by Equation 5.18.

$$Bias_{sc} = RSRP_s - RSRP_c \tag{5.18}$$

where $RSRP_s$ as the RSRP of subscriber from serving BS, and the second strongest RSRP can be represented as $RSRP_c$.

Therefore, the remaining resources of MC will be increased by the larger bias because more traffic can be offloaded from MC, and MC can allocate more resources to its serving subscribers. After the CRE is applied, the offloaded subscribers might suffer from high interference in PC's expanded area because of the short distance between subscribers and MC. Meanwhile, the value of $Bias_{sc}$ can not be too large because the coverage of PC is limited by the transmission power. The downlink performance will even worse than before if PC cannot assign more resources to the offloaded subscribers.

5.4.2 System Process

The system process can be divided into four phases, monitoring, estimation, calculation and implementation, which is shown in Figure 5.7.

Monitoring In monitoring phase, followed the load model in section 5.2.4, the BS's load information will be reported to the system in a certain time period.

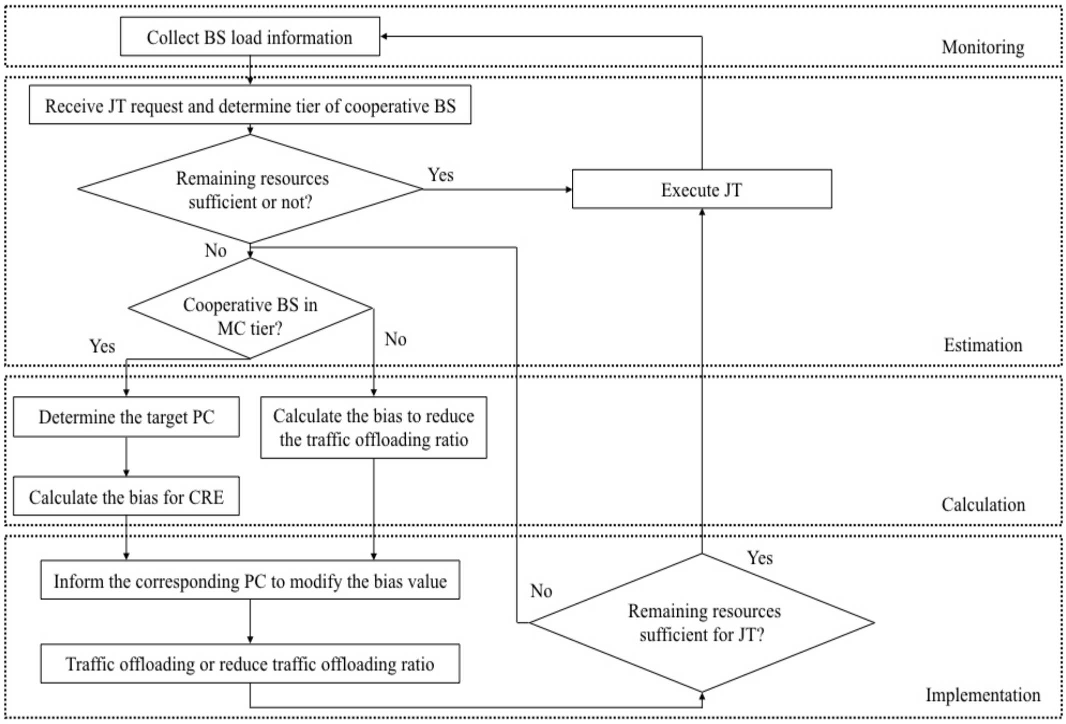


Fig. 5.7 System process.

Estimation The system receives JT request, the first step is to determine the cooperative BS, which is determined by the strongest received signal strength. After the decision is made, the remaining resources in cooperative BS will be estimated. The JT will be executed if the cooperative BS has sufficient resources, and the system will return to monitoring phase after the JT is finished. It the remaining resources is insufficient, the system will move to the next procedure.

Calculation The first step in calculation phase depends on the BS-tier of cooperative BS, and the objective of this phase is to inform the corresponding BS to modify the bias to achieve goal of load reduction.

• Cooperative BS in MC tier: The bias for CRE will be calculated after the system determines the target PC.

• Cooperative BS in PC tier: Cooperative PC will determine the offloaded traffic from MC, and the new bias will be calculated to reduce the offload ratio.

Implementation After the corresponding BS receive the bias from the system, the new cell association will be executed in target BS to reduce the load. The JT will be performed normally to improve cell-edge subscriber downlink performance, otherwise the process will return to estimation phase and follow the procdure to further reduce the load of cooperative BS until cooperative BS has sufficient resources for JT..

5.4.3 Problem Formulation

It is assumed that a source MC suffer from traffic congestion, and up to P target PCs are deployed within MC coverage. According to the purpose of CRE, the traffic will be offload from MC to PCs. Therefore, the system will equalise the traffic among P target PCs, where $\{1,2,3,...p\} \in P$. The traffic offloaded problem can be formulated as followed

$$min \quad f(x) = \left(\frac{\sum_{i=1}^{I} R_{im}X_{im}}{R_{tm}} - \frac{\sum_{i=1}^{I} R_{ip}X_{ip} + \sum_{i=1}^{I} R_{im}X_{im}}{\sum_{p=1}^{P} R_{tp} + R_{tm}}\right)^2 \tag{5.19}$$

$$st. \begin{cases} P_{im} \geqslant P_i, P_{ip} \geqslant P_i, \forall_m \in M, \forall_p \in P, \forall_i \in I \\ \sum_{i=1}^{I} R_{im}X_{im} \leqslant R_{tm}, \forall_m \in M, \forall_i \in I \\ \sum_{i=1}^{I} R_{ip}X_{ip} \leqslant R_{tp}, \forall_p \in P, \forall_i \in I \end{cases} \tag{5.20}$$

where P_{im} and P_{ip} represent the received signal strength of subscriber i when it connects to MC or PC, and P_i represents the minimum received signal strength from BS. The R_{im} and R_{ip} represent the allocated resources which is occupied by the subscriber i with corresponding serving BS. X_{im} and X_{ip} will equal to 1 if subscriber i is connected to respective BS, otherwise

0. The equations $\sum_{i=1}^{I} R_{im}X_{im} \leqslant R_{tm}$ and $\sum_{i=1}^{I} R_{ip}X_{ip} \leqslant R_{tp}$ represent the total allocated resources cannot be larger than the total resources of MC and PC.

5.4.4 NP-Hard Proof

In [125], [126] and [127], researches state that the partition optimisation problem is proved as a NP-hard problem in optimisation. The partition optimisation can be considered as a problem that n items have different value $v_1, v_2, v_3 ... v_n$ and different size $s_1, s_2, s_3 ... s_n$, and all items shall be packed into M bins with the same size S_{mb} that each of the bins shall have the same value without exceeding the size of the bins, the partition problem can be summarised as follows.

$$f(x) = \sum_{n=1}^{n} x_{n1}v_n = \sum_{n=1}^{n} x_{n2}v_n = \sum_{n=1}^{n} x_{n3}v_n ... = \sum_{n=1}^{n} x_{nm}v_n \tag{5.21}$$

$$st. \begin{cases} f(x) \geqslant 0, \\ \sum_{n=1}^{n} x_{nm}v_n \leqslant S_{mb} \\ \sum_{n=1}^{n} x_{nm} = 1, \end{cases} \tag{5.22}$$

This is the same optimisation problem that BSs have the fixed total resources, and they shall allocate the resources to the subscribers to provide service, but the total allocated resources from the serving subscribers can't exceed the total resources of one BS. However, the partition problem will become more complicated in cellular network since the allocated resources of one user will be different if the serving BS is changed. Therefore, the v value is different when the item n is put into different bin m, and the notation shall be changed to $v_{1m}, v_{2m}, v_{3m} ... v_{nm}$. The value of v_{21} is different with v_{22} since the path loss and interference will be changed if the serving BS is changed from 1 to 2, and the required resource will be changed as well. It can be seen that there will be nxm matrix to contain the v value. The partition problem can be summarised as follows. Therefore, the above optimisation problem is NP-hard.

$$f(x) = \sum_{n=1}^{n} x_{n1} v_{n1} = \sum_{n=1}^{n} x_{n2} v_{n2} = \sum_{n=1}^{n} x_{n3} v_{n3} \ldots = \sum_{n=1}^{n} x_{nm} v_{nm} \tag{5.23}$$

$$st. \begin{cases} f(x) \geqslant 0, \\ \sum_{n=1}^{n} x_{nm} v_{nm} \leqslant S_{mb} \\ \sum_{n=1}^{n} x_{nm} = 1, \end{cases} \tag{5.24}$$

5.4.5 Flow Water Algorithm

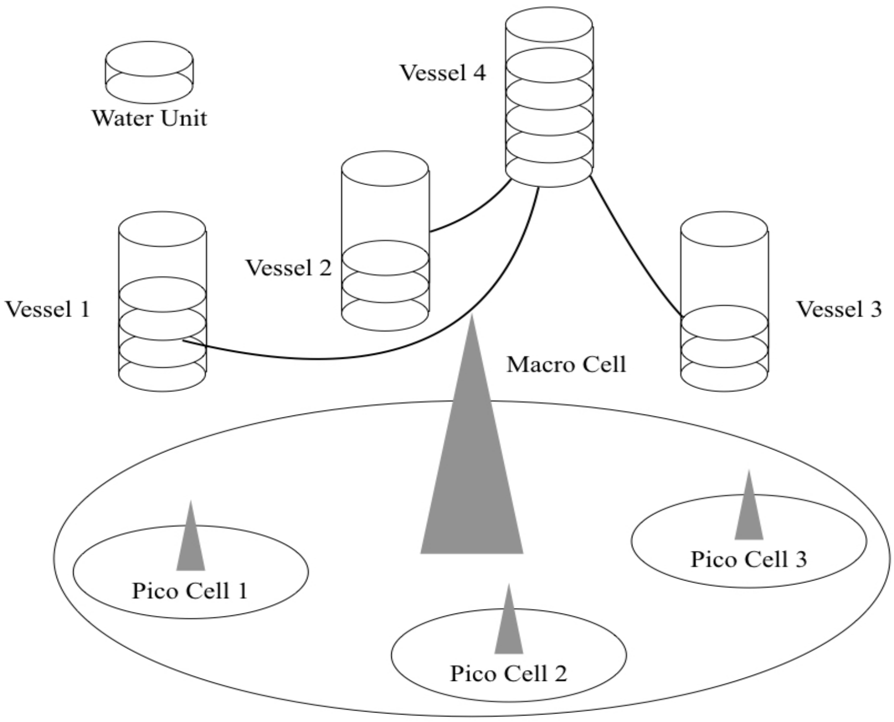


Fig. 5.8 Flow water concept.

The flow water algorithm is applied to tackle the above problem [129]. It is assumed that there are P vessels, and each vessel is connected by the pipes, which is shown in Figure 5.8. The central vessel has the largest volume of water, and vessel can transfer the water to the remaining vessels via the pipe when the controller is switched on, The valve will monitor the water pressure of each vessel, the water can flow to other low-pressure vessels. During the process, the system will switch on the lowest water pressure vessel's valve to receive the central vessel's water unit, and only one valve is allowed to be switched on. Only one water unit can flow to the target vessel each time the valve is switched on.

This concept is applied to the above problem. The central vessel represents the MC that traffic should redistribute to the PCs, the MC is shown in the figure as vessels 4. The side vessels, which is shown in the figure as vessels 1, 2 and 3, it represents the PCs within MC coverage, and the vessels are connected to the central vessel. The volume of the vessel represents the total resources of the BS, and the water units within the vessel represent the occupied resources of subscriber. The load of the BS is represented by the water pressure of the vessel, and the water within the vessel can only flow to the remaining vessels due to the water pressure.

Flow Water Algorithm Procedure The flow water algorithm is applied to tackle the traffic offloading problem. The objective of applying CRE is to alleviate the traffic congestion in MC, avoiding the overall network throughput degradation when JT is executed. The procedure is shown in the following Figure 5.9.

The system should collect the remaining resources of each BS, the algorithm will be executed if the cooperative BS cannot provide sufficient resources for JT. Firstly, the system will determine the target PC based on the current load of PC. After the target PC is determined, the MC will list the subscribers with descending order where the descending order is determined by the received signal strength from the target PC. Based on the list, the bias for the first subscriber will be calculated to offload traffic from MC to PC. After the traffic is offloaded to the target PC, the system will estimate the remaining resources of cooperative BS. If the remaining resources are not enough for JT, the algorithm will repeat the process until the JT can be executed.

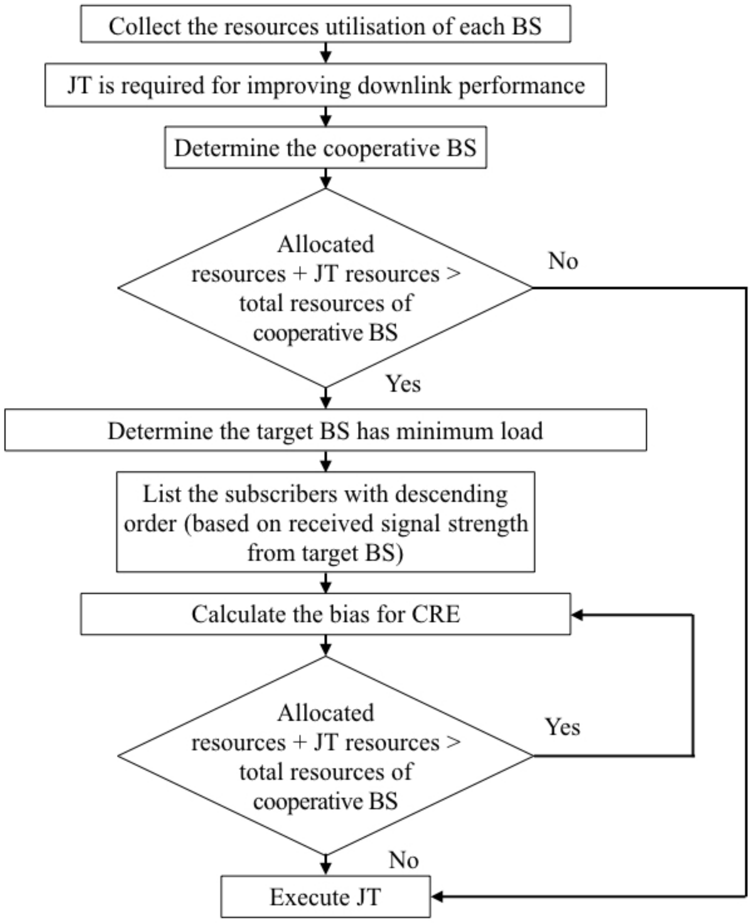


Fig. 5.9 Flow water algorithm procedure for joint transmission.

5.5 Simulation of M-P

5.5.1 Scenario

Let's consider the scenario that one subscriber appears in MC 1 edge, and it will move to MC 3 centre, JT should be applied to improve the downlink performance. For the simplicity of analysis, it is assumed that the remaining subscribers will not move during the simulation. The system chooses adjacent MC 3 as cooperative BS, but the traffic congestion is occurred in MC 3, and the cooperative BS cannot provide sufficient resources for JT. For example, if the subscriber require 4 Resource Blocks (RBs) from the cooperative BS, but the remaining

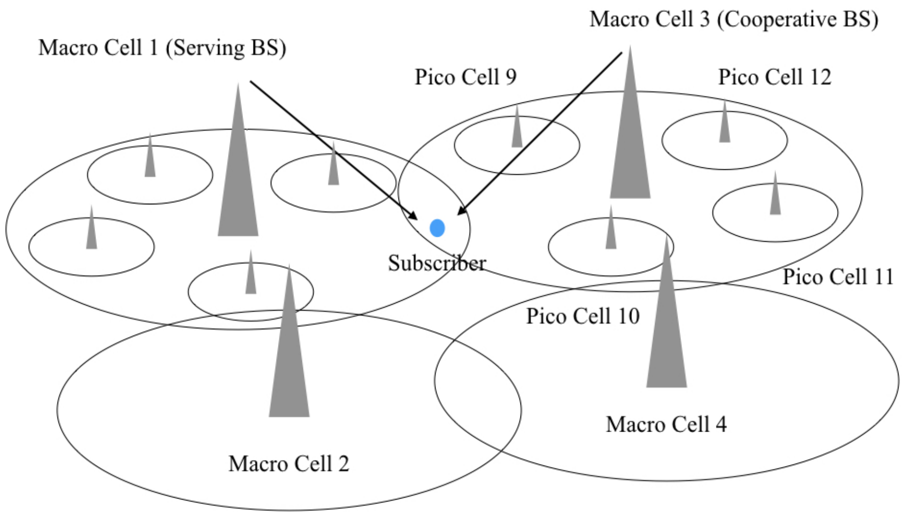


Fig. 5.10 Scenario of traffic offloading from macro cell to pico cells.

RBs are less than this value, it indicates that the remaining resources in cooperative BS is insufficient. However, 4 PCs are deployed within each MC coverage, CRE should be applied in low-load PC to offload the MC's traffic. The scenario is shown in Figure 5.10. By adjusting the bias in CRE, coverage of the PC will be expanded, and the traffic is offloaded to the PCs.

5.5.2 Offloading Traffic

The offloaded traffic ratio and the calculated bias in each iteration will be shown in Table 5.2.

According to the table, it shows that the traffic is offloaded from MC 3 to its coverage's PCs 9, 10, 11 and 12. After multiple times CRE are applied in different PC, the load of MC 3 is reduced from 99% to 64%. Traffic is offloaded to PCs 9, 10 and 12. Load of PC 9 is increased 8%, 22% traffic increment in PC 10 and 4% traffic increment in PC 12. Normally, bias will be a small number when the CRE is firstly applied in one PC. According to Table

Table 5.2 Offloading traffic from macro cell to pico cells.

Load of Base Stations						
CRE	Bias(BS)	Macro Cell 3	Pico Cell 9	Pico Cell 10	Pico Cell 11	Pico Cell 12
0	0(Original)	99%	33%	22%	59%	51%
1	2.20 (10)	91%	33%	30%	59%	51%
2	4.02 (10)	87%	33%	35%	59%	51%
3	0.39 (9)	75%	35%	35%	59%	51%
4	0.71 (9)	74%	36%	35%	59%	51%
5	10.3 (10)	67%	36%	47%	59%	51%
6	1.1 (9)	64%	37%	47%	59%	51%

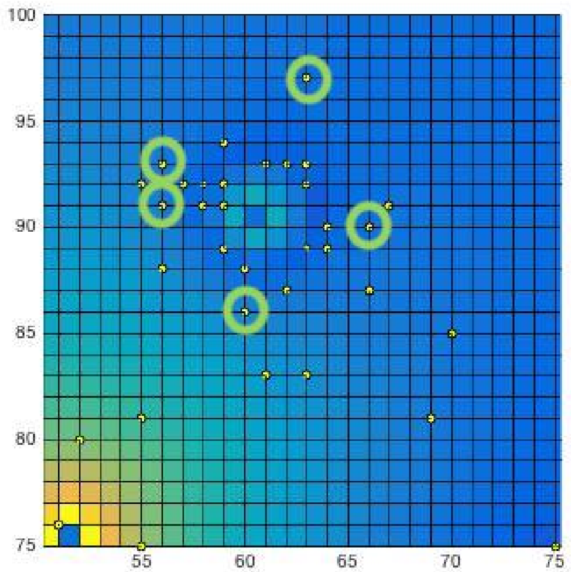


Fig. 5.11 Offload subscribers (First CRE).

5.2, the bias is 2.01 in PC 10 with first CRE, and the value is increased 4.02 in the second CRE, and the value of bias is 10.3 in the third CRE. Because the subscriber will be closer to the MC 3 if multiple times CRE are applied in the same PC, and a larger bias is needed to let subscriber chooses PC as the serving BS. This is also occur in PC 9 when the system execute CRE three times in the same PC, the bias is 0.39, 0.71 and 1.1.

The first CRE is used to demonstrate the offloading process, the load of MC 3 will be offloaded to PC10. The load of the PC is the same with the water pressure in the algorithm. Based on the algorithm, it begins to offload the traffic to the lowest-load PC first, which is PC 10 (location is (60, 95) in Figure 5.11) in the first CRE. The nearby yellow dots represent the subscribers which is served by the MC 3. The system will list the serving subscribers with descending order and determine the subscriber on the top of the list. According to the location of subscriber, two distance will be calculated, distance between subscriber and MC 3, distance between subscriber and PC 10. The bias will be given based on the bias calculation equation, and target PC will receive new bias and offload the traffic from MC. The process is the same with the water transfer from high water pressure vessel to low water pressure vessel. According to Table 5.2, the bias for the first CRE is 2.2. The green circles within Figure 5.11 shows the offload subscribers when the system choose PC 10 as target BS in the first CRE. The green circles is served by the MC 3 before CRE, serving BS is changed to PC 10 after PC 10 utilise the new bias. After the subscribers are served by new PCs, it represents the water is transferred to the low water pressure vessel.

5.5.3 Subscriber and System Throughput

The Figure 5.12 shows the throughput of subscriber with JT, JT gain and subscriber throughput during the whole simulation. The throughput of subscriber with JT is considered as the sum of the subscriber's achieved data rate in serving BS and cooperative BS, and the JT gain is considered as the throughput from cooperative BS only. The subscriber throughput in this part is considered as the throughput from serving BS.

According to the Figure 5.12, the blue line represents the throughput of subscriber with JT that it contains the throughput from serving BS and cooperative BS. Therefore, the blue line is on top of the orange line because the throughput of subscriber will not only come from serving BS but also the cooperative BS. The yellow line represents the throughput of subscriber without JT, when subscriber is moving toward the edge of the serving BS, the subscriber will suffer from higher path loss from serving BS and higher interference from the remaining BS, and it causes the subscriber will receive weaker SINR, causing the decreased throughput. Meanwhile, orange line represents the JT gain from cooperative BS, but the gain is limited at the beginning because of the path loss of MC 3 is high. However, the gain will raise due to the subscriber is moving toward the MC 3 centre, and the JT gain will larger than the subscriber's throughput loss after the location 2, and it will cause the throughput of subscriber with JT is increased after location 2, which is shown in the location 2 of blue line.

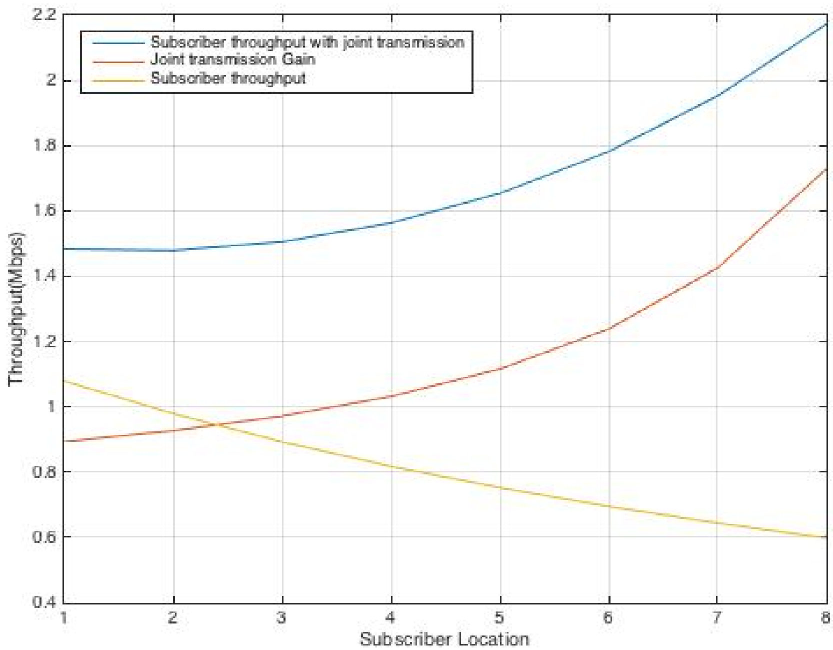


Fig. 5.12 Moving subscriber throughput.

The Figure 5.13 shows the total system throughput during the subscriber movement with and without JT. According to the figure, it can be seen that the subscriber throughput has a slightly decreased, which is shown at location 2 in Figure 5.13 because the throughput loss from serving BS is slightly larger than the JT gain from cooperative BS. This trend is followed the trend in location 2 of Figure 5.12. Because all the subscriber is assumed stable in this simulation, and the only moving subscriber in this simulation is the subscriber who requires JT. Meanwhile, the Figure 5.13 also shows that the throughput of the system is not influenced by the JT because the MC's traffic is offloaded to the PCs, MC 3 can provide sufficient resources to JT. Moreover, the remaining subscribers are assumed to be stable during the simulation, the throughput changed mainly comes from the serving BS and cooperative BS. Therefore, the total system throughput has a similar increment with the JT gain. During the whole simulation, the subscriber throughput with JT can obtain 26% average gain compared the subscriber's throughput without JT. Based on the simulation, it can be seen that the traffic can be sufficiently offloaded to PCs when the bias is added into PCs. Due to the sufficient remaining resources for JT, the system throughput will not be influenced, the moving subscriber can obtain optimal downlink.

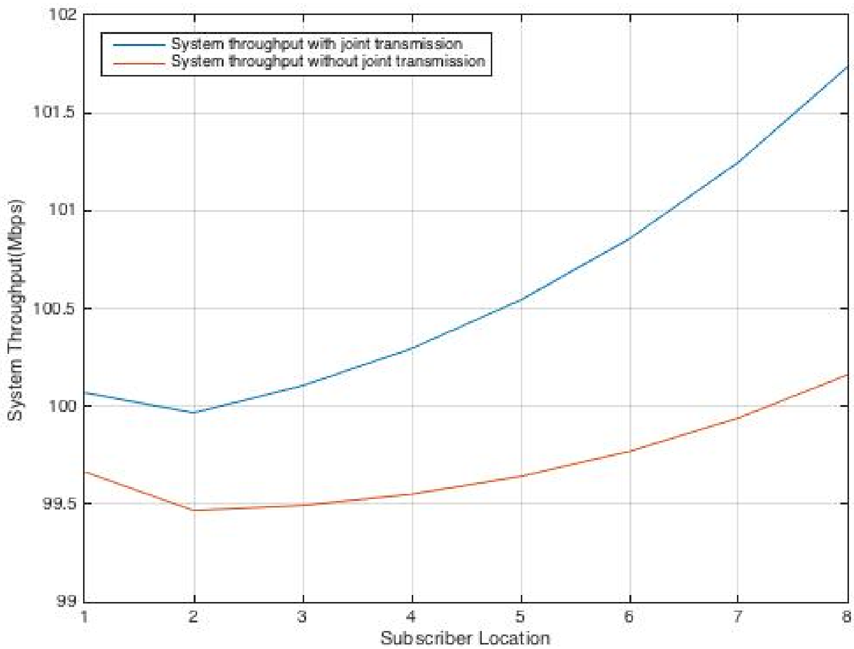


Fig. 5.13 System throughput.

5.6 Simulation of P-M

5.6.1 Scenario

Let's consider the scenario that JT is executed between MC 3 and PC 11, but cooperative PC is suffering traffic congestion due to the offloaded traffic from MC 3. Subscriber appear in MC 1 edge and move toward MC 3 centre, the scenario is shown in Figure 5.14. During the movement, JT is applied to improve the downlink performance. For the simplicity of the analysis, it is assumed that the remaining subscribers will be stable during the simulation. There are 4 PCs deployed within MC 3 coverage. The heavy-load PC 11 will reduce the offloaded traffic ratio to provide resources for JT. By adjusting the *Bias* in CRE, the coverage of cooperative PC and offloaded traffic ratio will be reduced.

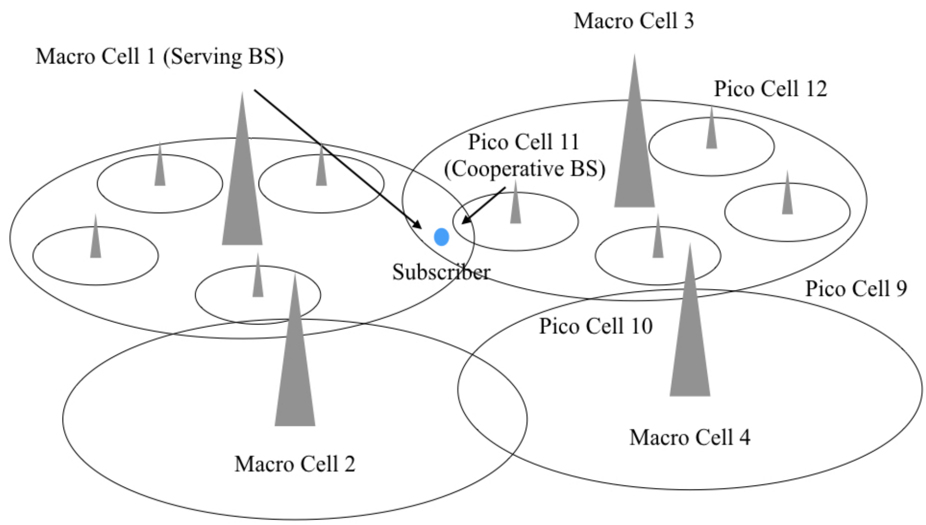


Fig. 5.14 Scenario of reducing traffic offloading from macro cell.

5.6.2 Reduced Traffic of Cooperative Pico Cell

This section will show the traffic reduction when the system reduce the bias in heavy-load cooperative PC. Table 5.3 shows the reduced bias for reducing the offloaded traffic from MC 3. The original load in PC 11 is 99% in the simulation, and this high load BS can be considered as the high water pressure vessel in the algorithm. The load of MC 3 is 74%, which is lower than PC11. Therefore, this BS and the remaining low-load BSs can be considered as the low water pressure vessel in the algorithm. In the first bias calculation, 0.7 bias value in PC 11 can be obtained. After the bias is utilised in cell selection, some subscribers are connected to MC 3, the load of MC 3 increases to 78%, and the load of PC 11 is reduced to 95%. This step can be considered as the water transfer from high pressure vessel to low pressure vessel in the algorithm. When the system calculate second bias value, it will repeat the same process. After the second bias calculation, 1.5 bias reduction in the same PC, and another 8% load is transferred back to MC 3. The process will continues when the system detect the cooperative BS has sufficient resources for the JT.

Table 5.3 Reduced offloaded traffic from macro cell.

	Load of Base Stations				
Reduced Bias(BS)	Macro Cell 3	Pico Cell 9	Pico Cell 10	Pico Cell 11	Pico Cell 12
0 (Original)	74%	54%	28%	99%	51%
0.7 (11)	78%	54%	28%	95%	51%
1.5 (11)	86%	54%	28%	87%	51%

5.6.3 Subscriber and System Throughput

Figure 5.15 shows the subscriber throughput with JT, JT gain from cooperative BS and the subscriber throughput without JT. At the beginning of the simulation, the throughput of moving subscriber with JT is decreased because the serving BS provides a weak signal when subscriber is moving toward the MC 1 edge. Meanwhile, PC 11 is the cooperative

BS, and it cannot provide significant gain in downlink because the transmission power of PC 11 is small. High path loss cause the received signal strength of PC 11 is not optimal in subscriber's side, and cooperative BS brings little effort during the subscriber movement, which is shown in the orange line of Figure 5.15 from 1 to 4. Therefore, even the subscriber throughput is higher than before, the trend is decreased. However, the gain from cooperative BS begins to increase when subscriber is moving toward PC 11 coverage, which is shown is the Figure 5.15 from 4 to 7, and it can be seen that the subscriber throughput with JT is increased during this period, which is shown in blue line in Figure 5.15 from location 4 to 7. In Figure 5.15, orange line shows that the gain is increased significantly in the last three locations due to the subscriber is moving very close to the PC 11 centre, the received signal strength is optimal, it can bring significant gain in JT, and the significant gain also cause the subscriber's throughput is raised from 5 to 7 in Figure 5.15.

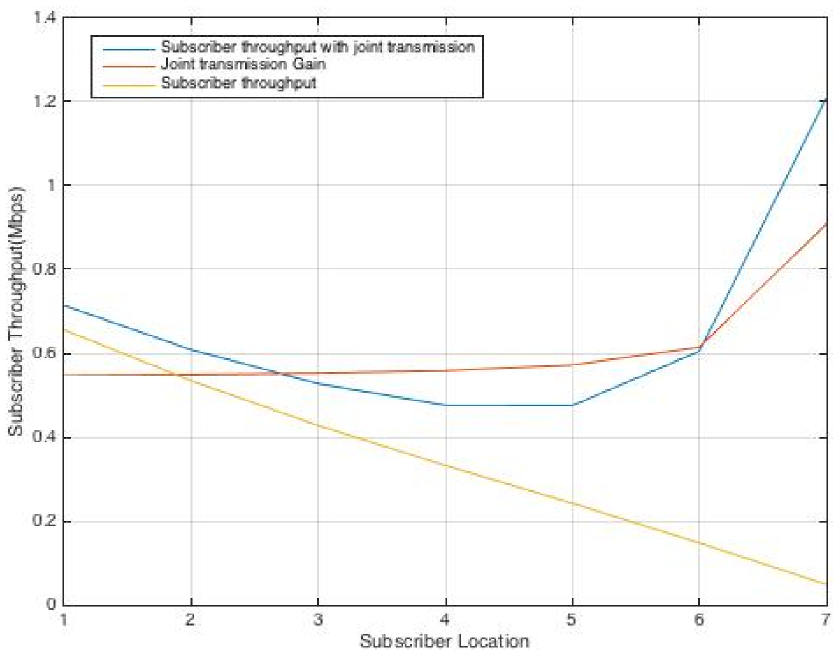


Fig. 5.15 Subscriber throughput.

The Figure 5.16 shows the total system throughput during the simulation with and without JT. Because there is only one moving subscriber in this simulation, the remaining subscribers

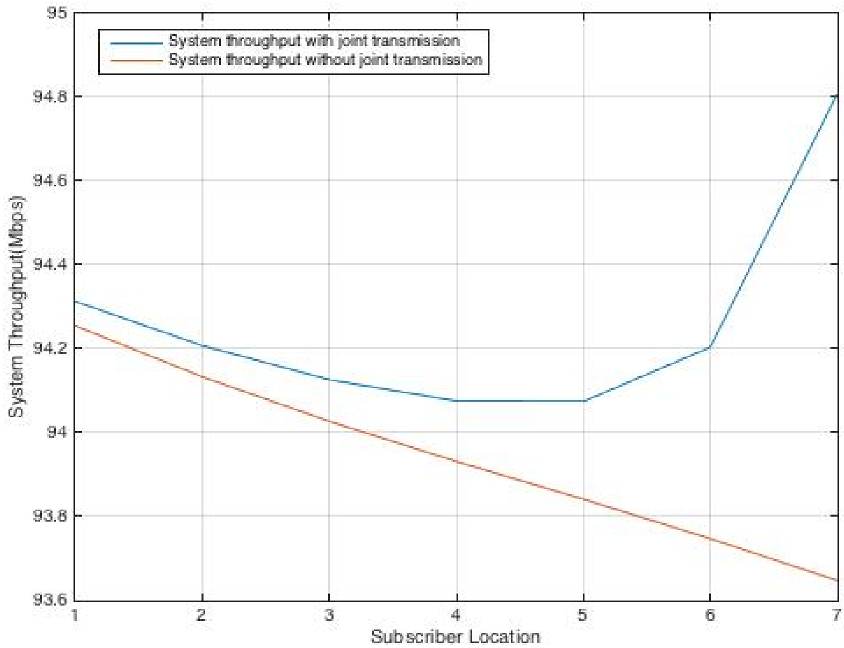


Fig. 5.16 System throughput.

will be considered as stable. Therefore, the change in system throughput is caused by the moving subscriber. Due to the subscriber is moving out the coverage of MC 1 at the beginning, the throughput will be decreased because the weak signal from serving BS and cooperative BS, which is shown in Figure 5.16 from location 1 to 4 . However, the loss from serving BS will become insignificant when the subscriber can obtain higher gain from cooperative BS, which is shown in the Figure 5.16 from 4 to 7. Compared with the moving subscriber without JT, it can be seen that the throughput difference is obvious, even both lines have the decreased trend, the subscriber can still obtain higher throughput than without JT. The average improvement during the whole simulation is 24%.

The simulation result shows that the JT will not be influenced by the insufficient resources issue in cooperative BS. When the JT is triggered, the algorithm will calculate the bias and inform the target BS to offload the subscribers in PC 11 (high-load BS) to the MC 3 (low-load BS) to provide resources for the JT. Beside that, it can be seen that the executed time of JT is also important. The subscriber can only obtain the significant gain from cooperative BS when it can receive a optimal signal. If the JT is executed too early, the signal from JT is weak, and the gain from cooperative is merely nothing, it will waste the resources to obtain such little throughput gain. In the simulation, it only needs to trigger JT when the subscriber is moving into location 4.

In the sections 5.5 and 5.6, both scenarios shows that the algorithm can successfully offload the traffic from high-load BS to nearby low-load BSs. Meanwhile, the remaining resources are used to participate the JT without sacrificing other subscribers' resources. Even the throughput goes down in some of the time slots in the Figure 5.13 and Figure 5.16, but the decrement is caused by the increasing path loss from serving BS not from the resources issue.

5.7 Conclusions

In this chapter, JT resources shortage problem is investigated, different JT gain is shown when different cooperative BS is chosen, and the consequence of high-load cooperative is discussed. Due to the same amount of occupied resources in multiple BSs, the throughput loss might occur if high-load BS participates the JT because cooperative BS will utilise other subscriber's resources to execute JT. According to the simulation, it can be seen that if the gain from JT cannot be larger than the original subscriber's throughput, the throughput loss will occur. The simulation also shows that subscriber can still obtain gain from JT if remote BS is chosen as cooperative BS. However, this cooperative BS selection shall be avoided because of the low spectrum efficiency.

CRE is proposed to tackle the traffic offloading problem, the bias will be added when subscriber make the decision of connected BS, larger value of bias can offload more traffic from MC. Therefore, the bias can be modified when the system detect the high-load MC is chosen as cooperative BS. The problem is formulated as an optimisation problem, but it will become a NP-hard problem with large number of PC and subscriber. The flow water algorithm is applied to provide the solution, and the modified algorithm will consider the load status and remaining resources. MC will offload part of the traffic to the lowest-load PC in each iteration, and the bias will be calculated after the target PC is determined. The load of MC can be significantly reduced many iterations. However, if the resources problem occurs in PC tier, the PC should reduce the traffic offloaded ratio from MC. The simulation shows that the system throughput will not be influenced by the JT after the resources shortage problem is tackled. To conclude, instead of choosing remote BS to avoid the system throughput loss, CRE can sufficiently reduce the cooperative BS's load, and the system can choose the closest BS to execute JT.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, reducing energy consumption and spectrum efficiency are investigated. The system switch off BSs to reduce the energy consumption of the network, and it utilises SA to tackle PCs combination problem that determines the best PCs combination to achieve the best EE. Second, it collects and analyse subscriber's data and use NBC to predict the traffic distribution. The prediction result is used to inform the corresponding BSs to prepare the incoming traffic, reducing the energy consumption of the network. At last part, it proposes FWA to solve the unbalanced traffic between MCs and PCs, tackling the problem of heavy load in cooperative BS. The JT can be executed without sacrificing the overall network performance after the LB.

Chapter 3 utilise SA algorithm to determine the best PCs combination to achieve optimal EE after the system determines the number of switch-off PC, and energy consumption can be reduced. The optimisation problem becomes a combinational problem that the system should choose a certain number of PCs from the total PC group. The cost function will be estimated in each iteration when SA chooses a new PCs combination. SA algorithm will accept the PCs combination when EE is better than the current combination, and SA will also accept the worse PCs combination with a certain probability to avoid the local maximum value. However, the algorithm will begin to reject worst PCs combination with increasing iteration times. The simulation results show that when the system is choosing the PCs to serve the existing subscribers, it will consider the location of the subscribers and the required downlink speed. Therefore, the best EE if the original light-load PCs are switched off because the location of PCs is different in each PCs combination, the subscriber's SINR might not be identical, and the consequence is the difference in EE. The multiple simulations show that the average 17.06% EE improvement can be obtained. After the traffic is transferred to neighbouring PCs, the resources utilisation is increased. The simulation result also shows that the EE can be improved when resources utilisation is increased given the same number of PCs are switched off. All simulation results prove that switching off some light-load PCs can reduce the energy consumption of the network, improving EE and resources utilisation.

Chapter 4 proposes the NBC to analyse and utilise subscribers information to predict the traffic distribution in CN, and power of PC will be modified based on the prediction result. The features of walking speed, downlink, remaining time and direction are collected. After subscribers' data is filtered by the system, the remaining subscribers' data will move to analysis phase, and correlation between subscribers features and PCs can be determined, NBC will utilise the analysis result to predict traffic. In the simulations, it shows that walking speed, downlink and remaining time can help the system to distinguish the difference in distance. However, these features cannot further improve the prediction accuracy because the value of these features are similar in closed PCs group. Therefore, direction feature is added into NBC to improve prediction accuracy. After the direction is added, the system can distinguish the difference even some PCs are in the same group. The multiple simulations show that average 90.8% prediction accuracy can be obtained when all features are utilised to traffic prediction. After traffic prediction can be obtained, the system can foresee the incoming traffic, the power will be modified according to the load adaptive power model. The power of PC will be increased proportional to the predicted load of PCs to match the growing traffic demand. When the traffic demand is continues growing, the system will further inform the corresponding PC to increase their transmission power. According to the simulation results, it shows that it can significantly reduce up to 67% the network energy consumption and achieve 4.7% EE improvement. The simulations prove that the proposed ML method can be used to predict the traffic distribution in CN, and load adaptive power model can cooperate well with prediction result.

Chapter 5 investigates the JT resources shortage problem, and the influence of executing JT in high-load cooperative BS is shown. JT requires multiple BSs to transmit signal to one subscriber to improve the downlink performance, but it might cause the throughput loss if high-load BS involve JT because cooperative BS will utilise other subscribers' resources to execute the JT if resources shortage problem is occurred. The simulations show that the throughput loss will occur if the gain from JT can not be larger than the original subscriber's throughput. However, subscriber can obtain little gain from JT if the system choose remote BS as cooperative BS. To tackle this issue, CRE is applied to offload traffic between MC and PCs, resources problem in high-load cooperative BS can be tackled. The FWA helps the traffic transaction that it solves the resources shortage issue in cooperative. The simulations shows that the traffic can be offloaded via modifying the CRE bias, and the subscriber can obtain 26% average gain in downlink, and 24% downlink gain in another scenario. The work in chapter 5 proves that the proposed algorithm can reduce the load of cooperative BS, and the downlink gain can be obtained without scarifying overall network throughput.

6.2 Future Work

The cell switch-off strategy is proposed to reduce the energy consumption of CN, optimisation is formulated and solved by the SA, achieving optimal EE. NBC is utilised to predict the traffic distribution, the subscriber features are collected and analysed to predict the subscriber destination. The power of PC is modified based on the traffic prediction result. CRE is utilised to offload traffic, tackling the resources shortage problem in cooperative BS. However, potential research directions are summarised in the following parts.

In Chapter 3, the SA is applied to determine the best PCs combination with small scale PC and subscriber number. It can be seen that the increasing number of PCs will be deployed in 5G, and the computation complexity will be significantly increased. Moreover, the system will face the challenge in defining the PC group because of high-density PC deployment will let PC group's contour become fuzzy. The power of BSs can be one of the optimisation parameters in future work, and the BSs can be operated in different power.

In Chapter 4, NBC is proposed to predict the distribution based on the collected subscriber data. Because the data is filtered by the system, the classifier can only utilise strong pattern data to predict the traffic distribution. Further improvement should be made to fully utilise the remaining subscriber's data, discovering the pattern within random data. Moreover, the current chosen features cannot distinguish the PC's difference in high-density PC deployment, features chosen should be deeper considered. However, multiple features in NBC will increase the time in computation. Therefore, the classifier should have the ability to choose appropriate features to predict traffic in different timing.

In Chapter 5, The CRE technic is applied to tackle the resources shortage problem in cooperative BS. However, if all MCs and PCs suffer from traffic congestion, the surrounding BSs cannot offload the traffic from cooperative BS, JT might cause impact on the overall system throughput. Advanced traffic offloading method shall be investigated in the future, the JT can be executed at any time and any BS. Meanwhile, the interference from other cooperative BSs will become a issue in 5G due to multiple antennas are deployed in one site. The interference mitigation in JT shall be further investigated.

The extension of this thesis, it will be the combining the prediction technic in Chapter 3 and JT in Chapter 5. Due to the occupied resources in serving and cooperative BS, the system cannot apply JT to one subscriber for a long time, and JT should be executed only the current service cannot be fulfilled. The potential research area will be the subscriber movement prediction to give the system a perfect timing to trigger the JT. Therefore, the JT's execution time can be reduced.

References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update 2016-2021," White paper, Mar. 2017.

[2] D. Calin, H. Claussen, and H. Uzunalioglu, "On femto deployment architectures and macrocell offloading benefits in joint macro-femto deployments," *IEEE Commun. Mag.,* vol. 48, no. 1, pp. 26-32, 2010.

[3] J. Zhang and G. de la Roche, *Femtocells; Technologies and Deployment.* John Wiley & Sons Ltd., 2009.

[4] R. Hu and Y. Qian, "An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems," *IEEE Commun. Mag.,* vol. 52, no. 5, pp. 94-101, May 2014.

[5] R. Bassoli, M. Di Renzo, F. Granelli, "Analytical energy-efficient planning of 5G cloud radio access network," in *Proc. 2017 IEEE International Conference on Communications (ICC)*, Paris, France, 21-25 May 2017, pp. 1-4.

[6] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M.A. Imran, D. Sabella, M.J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun. Mag.,* vol. 18, no. 5, pp. 40-49, Oct. 2011.

[7] S. Fortes, A. Aguilar-García, R. Barco, F. B. Barba, J. A. Fernández-Luque, and A. Fernández-Durán, “Management architecture for location-aware self-organizing LTE/LTE-A small cell networks,” *IEEE Commun. Mag.,* vol. 53, no. 1, pp. 294-302, Jan. 2015.

[8] O.G. Aliu et al., “A Survey of Self Organisation in Future Cellular Networks,” *IEEE Commun. Surveys & Tutorials,* vol. 15, no. 1, Jan. 2013, pp. 336-61.

[9] J.T. Louhi, “Energy Efficiency of Modern Cellular Base Stations,” *IEEE INTELEC ’07,* Rome, Italy, Sept. 2007.

[10] F. Richter, A. J. Febske, and G. P. Fettweis, “Energy Efficiency Aspects of Base Station Deployment Strategies in Cellular Networks,” Proc. *IEEE VTC ’09-Fall,* Anchorage, AK, Sept. 2009.

[11] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, “Green cellular networks: A survey, some research issues and challenges,” *IEEE Commun. Surveys & Tutorials,* vol. 13, no. 4, pp. 524-540, Fourth Quarter 2011.

[12] M. Deruyck, W. Joseph, B. Lannoo, D. Colle, and L. Martens, “Designing energy-efficient wireless access networks: LTE and LTE- Advanced,” *IEEE Internet Comput.,* vol. 17, no. 5, pp. 39-45, Sep./Oct. 2013.

[13] L.M. del Apio et al., “Energy Efficiency and Performance in Mobile Networks Deployments with Femtocells,” *Proc. IEEE Int’l Symp. Personal Indoor and Mobile Radio Comm.,* IEEE, 2011, pp. 107-111.

[14] L.M. Correia et al., “Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks,” *IEEE Commun. Mag.,* Nov. 2010, pp. 66-72.

[15] A. Jahid, A.B. Shams, and M.F. Hossain, “Energy efficiency of JT comp based green powered lte-a cellular networks,” *IEEE Conference on Wireless Communications,*

Signal Processing and Networking (WiSPNET), Chennai, India, to appear, Mar. 2017, pp. 1-6.

[16] D.A. Danecha, K. G. Bhuva and S. K. Dastoor, “Energy efficiency analysis for LTE-A heterogeneous cellular network,” *ICTACT journal on communication technology,* Sep. 2017, vol. 08.

[17] M. Prasad, “Energy efficiency in LTE-A network,” *2017 International Conference on Technical Advancements in Computers and Communications (ICTACC),* Apr. 2017.

[18] A.M. Hamed and R.K. Rao, “Energy and spectral efficiency in cellular networks considering fading, path loss, and interference,” in *Proc. of the 30th IEEE Canadian Conference on Electrical and Computer Engineering, CCECE 2017,* Canada, May 2017.

[19] F.R. Arvaje, B.S. Ghahfarokhi, “A Spectrum efficient base station switching-off mechanism for green cellular networks,” *2017 7th International Conference on Computer and Knowledge Engineering (ICCKE)*, Oct. 2017.

[20] E. Dahlman et al., “3G Radio Access Evolution - HSPA and LTE for Mobile Broadband,” *IEICE Trans. Commun.,* vol. E92-B, no. 5, May 2009, pp.1432-1440.

[21] F. Alaca, A.B. Sediq, and H. Yanikomeroglu, “A genetic algorithm based cell switch-off scheme for energy saving in dense cell deployments,” in *Proc. IEEE GLOBECOM Workshop,* Anaheim, CA, USA, Dec. 2012, pp. 63-68.

[22] B. Debaillie, C. Desset, and F. Louagie, “A flexible and future-proof power model for cellular base stations,” in *Proc. IEEE VTC-Spring,* May 2015, pp. 1-6.

[23] Z. Niu, “TANGO: traffic-aware network planning and green operation,” *IEEE Wireless Commun. Mag.,* vol. 18, no. 5, pp. 25-29, Oct. 2011.

[24] I. Ashraf, F. Boccardi, and L. Ho, "Sleep mode techniques for small cell deployments," *IEEE Commun. Mag.,* vol. 49, no. 8, pp. 72-79, Aug. 2011.

[25] "Overview to LTE energy saving solutions to cell switch off/on," 3GPP TSG-RAN WG3 #66 R3-100162, Huawei, Jan. 2010.

[26] "LTE energy saving solution proposal following discussions at RAN3#65 bis," 3GPP TSG-RAN WG3 #66 R3-093103, Ericsson, Nov. 2009.

[27] J. Gong, S. Zhou, and Z. Niu, "A dynamic programming approach for base station sleeping in cellular networks," *IEICE Trans. Commun.,* vol. E95-B, no. 2, pp. 551-562, Feb. 2012.

[28] Y. E. Morabit, F. Mrabti and E. H. Abarkan, "Small cell switch off using genetic algorithm," *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP),* Fez, 2017, pp. 1-4.

[29] L. AlKanj, W. ElBeaino, A. M. ElHajj and Z. Dawy, "Optimized Joint Cell Planning and BS ON/OFF Switching for LTE Networks," *Wireless Communications and Mobile Computing 16(12),* 1537-1555, 2016.

[30] G. Mohammed, D. Soufiane, M. El Bekkaye and S. Abdelhafid, "Optimization Of Multi-Objective and Green LTE RNP Problem," *2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS),* Fez, Morocco, 2019, pp. 1-6.

[31] Z. Niu, Y. Wu, J. Gong and Z. Yang, "Cell zooming for cost-efficient green cellular networks," *IEEE Commun. Mag.,* vol.48, no.11, pp.74-79, Nov. 2010.

[32] Y. Gao, Y. Li, H. Yu, X. Wang, S. Gao, and P. Xue, "Energy efficient cooperative sleep control using small cell for wireless networks," *International Journal of Distributed Sensor Networks.*, 2015.

[33] S. Cai, L. Xiao, H. Yang, J. Wang, and S. Zhou, "A cross-layer optimization of the joint macro- and picocell deployment with sleep mode for green communications," in *Proc. 22nd WOCC,* May 2013, pp. 225-230.

[34] P. Dini, M. Miozzo, N. Bui, and N. Baldo, "A model to analyze the energy savings of base station sleep mode in LTE HetNets," in *Proc. IEEE GreenCom, IEEE iThings/CPSCom,* Aug. 2013, pp. 1375-1380.

[35] L. Saker, S.E. Elayoubi, R. Combes, and T. Chahed, "Optimal control of wake up mechanisms of femtocells in heterogeneous networks," *IEEE J. Sel. Areas Commun.,* vol. 30, no. 3, pp. 664-672, Apr. 2012.

[36] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning. Cambridge,* MA, USA: MIT Press, 2012.

[37] E. Alpaydin, *Introduction to Machine Learning.* MIT Press, 2004.

[38] D. Barber, *Bayesian Reasoning and Machine Learning,* Cambridge University Press, 2012.

[39] I. Androutsopoulos, J. Koutsias, K. V. Chandrinos, G. Paliouras, and C. D. Spyropoulos. "An evaluation of naive bayesian anti-spam filtering." in *Proc. of the 11th European Conference on Machine Learning,* pages 9-17, 2000.

[40] S. Rogers, M. Girolami, *A First Course in Machine Learning,* CRC Press, 2011.

[41] K.P. Murphy, "Naive Bayes classifiers," Technical Report, October, 2006,available:http: //www.cs.ubc.ca/murphyk/Teaching/CS340-Fall06/reading/NB.pdf.

[42] W.M. Bolstad, *Introduction to Bayesian Statistics.* Hoboken, NJ: Wiley, 2007.

[43] K.R. Koch, "Bayesian statistics and Monte Carlo methods." in *International Journal on Geomathematics.* vol. 9, pp. 117-143, Apr. 2018.

[44] I. Rish, "An empirical study of the naive Bayes classifier," in *Proc. IJCAI Workshop Empirical Methods AI,* 2001, pp. 41-46.

[45] S. Akoush and A. Sameh, "The use of Bayesian learning of neural networks for mobile user position prediction," in *Proc. 7th Int. Conf. Intell. Syst. Design Appl. (ISDA),* Oct. 2007, pp. 441-446.

[46] R. Barco, V. Wille, and L. Díez, "System for automated diagnosis in cellular networks based on performance indicators," *Eur. Trans. Telecommun.,* vol. 16, no. 5, pp. 399-409, 2005.

[47] R. Singh, M. Srinivasan, C. Murthy, "A learning based mobile user traffic characterization for efficient resource management in cellular networks." in *12th Annual IEEE Consumer Communications and Networking Conference (CCNC),* pp. 304-309, Jan. 2015.

[48] D. DelTesta, M. Danieletto, G.M. DiNunzio and M. Zorzi, "Estimating the Number of Receiving Nodes in 802.11 Networks via Machine Learning Techniques," in *2016 IEEE Global Communications Conference (GLOBECOM).* IEEE, Dec. 2016, pp. 1-7.

[49] X. Lin, Y. Yu, T. Qiu, C. Chen, Z. Mi, "A Novel Behavior-Based Prediction Dynamic Model for Mobility Management," in *9th IEEE International Conference Communication Software and Networks (ICCSN),* May 2017.

[50] N. A. Amirrudin, S. H. S. Ariffin, N. N. N. A. Malik, and N. E. Ghazali, "Mobility Prediction via Markov Model in LTE Femtocell," *International Journal of Computer Applications,* vol. 65, no. 18, pp. 40-44, 2013.

[51] I. Jeon, K. Lee, R. Jeong, "A prediction mobility model for delay tolerant networks," in *5th IEEE International Conference Consumer Electronics (ICCE),* Sep. 2015.

[52] A. Mohamed et al., “Mobility prediction for handover management in cellular networks with control/data separation,” in *Proc. IEEE Int. Conf. Commun. (ICC),* London, U.K., Jun. 2015, pp. 3939-3944.

[53] H. Si, Y. Wang, J. Yuan, and X. Shan, “Mobility prediction in cellular network using hidden Markov model,” in *Proc. 7th IEEE Consum. Commun. Netw. Conf., Las Vegas,* NV, USA, Jan. 2010, pp. 1-5.

[54] D. Stynes, K. N. Brown, and C. J. Sreenan, “A probabilistic approach to user mobility prediction for wireless services,” *2016 Int. Wirel. Commun. Mob. Comput. Conf.,* pp. 120-125, 2016.

[55] S. Almeida, J. Queijo, and L. M. Correia, “Spatial and temporal traffic distribution models for GSM,” in *Proc. IEEE VTC Fall, Amsterdam,* The Netherlands, Sep. 1999, pp. 131-135.

[56] S. Isaacman et al., “Human mobility modeling at metropolitan scales,” in *Proc. ACM MobiSys,* Low Wood Bay, Lake District, U.K., Jun. 2012, pp. 1-13.

[57] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, “Characterizing and modeling internet traffic dynamics of cellular devices,’ in *Proc. 2010 ACM SIGMETRICS*.

[58] X. Chen, F. Mériaux, and S. Valentin, “Predicting a user’s next cell with supervised learning based on channel states,” in *Proc. IEEE Int. Workshop SPAWC,* Jun. 2013, pp. 36-40.

[59] C. Yu et al., “Modeling user activity patterns for next-place prediction,” *IEEE Syst. J.,* vol. 11, no. 2, pp. 1060-1071, Jun. 2017.

[60] D.S. Wickramasuriya, C.A. Perumalla, K. Davaslioglu and R.D. Gitlin, “Base Station Prediction and Proactive Mobility Management in Virtual Cells using Recurrent Neural Networks,” in *IEEE Wireless and Microwave Technology Conference,* 2017.

[61] S. Zhang, S. Zhao, M. Yuan, J. Zeng, J. Yao, M. R. Lyu, and I. King, “Traffic prediction based power saving in cellular networks: A machine learning method,” in *Proc. of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,* Redondo Beach, CA, USA, Nov 2017, p. 29.

[62] D. Zhang, M. Chen, M. Guizani, H. Xiong, D. Zhang, “Mobility Prediction in Telecom Cloud Using Mobile Calls.” *IEEE Wirel. Commun.* 2014, 21, 26-32.

[63] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.P. Mayer, L. Thiele, and V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” *IEEE Commun. Mag.,* vol. 49, no. 2, pp. 102-111, Feb. 2011.

[64] 3GPP TR 36.819, “Coordinated Multi-Point Operation for LTE Physical Layer Aspects,” Release 11, V2.0.0, Sept. 2011.

[65] S. Sun, Q. Gao, Y. Peng, Y. Wang, and L. Song, “Interference management through CoMP in 3GPP LTE-advanced networks,” *IEEE Wireless Communications,* vol. 20, no. 1, pp. 59-66, 2013.

[66] D. Lee et al., “Coordinated multipoint transmission and reception in LTE-Advanced: Deployment scenarios and operational challenges,“ *IEEE Commun. Mag.,* vol. 50, no. 2, pp. 148-155, Feb. 2012.

[67] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, M. Tanno, “Coordinated multipoint transmission/reception techniques for LTE-advanced,” *IEEE Wireless Communications,* Vol. 17, Issue 3, pp 26-34, 2010.

[68] L. Poongup et al., “Interference Management in LTE Femtocell Systems Using Fractional Frequency Reuse,” *Proc. 12th Int’l. Conf. Advanced Commun. Tech.,* vol. 2, 7-10 Feb. 2010, pp. 1047-51.

[69] J. Li, N. Shroff, and E. Chong, "A Reduced-Power Channel Reuse Scheme for Wireless Packet Cellular Networks," *IEEE/ACM Trans. Net.,* vol. 7, no. 6, Dec. 1999, pp. 818-32.

[70] 3GPP R2-106897, "Introduction of Enhanced ICIC," 3GPP TSG RAN WG2 meeting #72.

[71] D.López-Pérez, I. Guvenc, G. delaRoche, M. Kountouris, T.Q.S. Quek and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Commun. Mag.,* vol. 18, no. 3, pp. 22-30, Jun. 2011.

[72] X.Q. Lu, S. Chen, W.A. Zhou and J.D. Song, "Analysis of Signaling Design in LTE-Advanced Coordinated Multipoint Transmission/Reception System," *IEEE Conference on Computer Science and Automation Engineering (CSAE),* Zhangjiajie, China. 2012.

[73] B. Mondal, E. Visotsky, T. A. Thomas, X. Wang, and A. Ghosh, "Performance of downlink CoMP in LTE under practical constraints," in *Proc. IEEE 23rd Int. Symp. PIMRC,* 2012, pp. 2049-2054.

[74] W. Lee, H. Lee, "Performance Evaluation of coordinated multi-point transmission and reception in indoor mobile communication systems, *J. Inf. Commun. Converg. Eng.,* 11(3): 167-172, 2013.

[75] A. Davydov et al., "Evaluation of Joint Transmission CoMP in C-RAN Based LTE-A HetNets with Large Coordination Areas," *2013 IEEE GLOBECOM Wksps.,* Dec. 2013, pp. 801-06.

[76] T. Okamawari, H. Hayashi,T. Fujii, "A Proposal on Network Control Architecture for CoMP JT with IP Network between eNBs," in *Proc. IEEE 75th Vehicular Technology Conference (VTC2012 Spring),* May 2012.

[77] C. Khirallah, D. Vukobratovic, and J. Thompson, “On energy efficiency of joint transmission coordinated multi-point in lte-advanced,” in *2012 International ITG Smart Antennas (WSA) Workshop,* 2012, pp. 54-61.

[78] P. Bhat et al., “LTE-Advanced: An Operator Perspective,” *IEEE Commun. Mag.,* vol. 50, no. 2, Feb. 2012, pp. 104-14.

[79] J. Liu, Y. Chang, Q. Pan, X. Zhang and D. Yang, “A novel transmission scheme and scheduling algorithm for CoMP SU-MIMO in LTE-A system,” *IEEE VTC2010-Spring,* pp. 1-5, May 2010.

[80] N. Abbas, T. Bonald, B. Sayrac, “How mobility impacts the performance of inter-cell coordination in cellular data networks,” in *IEEE Globecom,* San Diego, CA, USA, Dec. 2015.

[81] N. Abbas, T. Bonald, B. Sayrac, “Mobility-aware scheduler in CoMP systems,” in *2016 IEEE 27th Annual International Symposium Personal, Indoor, and Mobile Radio Communications (PIMRC),* Sep. 2016.

[82] W. Webb, “Wireless Communications: The Future,” John Wiley & Sons, 2007.

[83] P. Gandotra, R. Jha, S. Jain, “Green communication in next generation cellular networks: a survey,” in *IEEE Access,* vol. 5, no.99, pp. 11727-11758, 2017.

[84] J. Wu, Y. Zhang, M. Zukerman, and E. K. N. Yung, “Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey,” *IEEE Commun. Surveys Tuts.,* vol. 17, no. 2, pp. 803-826, May 2015.

[85] H. Holtkamp, G. Auer, S. Bazzi, and H. Haas, “Minimizing base station power consumption,” *IEEE J. Sel. Areas Commun.,* vol. 32, no. 2, pp. 297-306, Feb. 2014.

[86] 3GPP, "Potential Solutions for Energy Saving for E-UTRAN (Release 10)," Third-Generation Partnership Project, Sophia-Antipolis Cedex, France, TS 36.927 V2.0.0, 2011.

[87] M. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *Proc. Green Communications, Dresden,* Germany, June 2009, pp. 1-5.

[88] W. Guo and T. O'Farrell, "Green cellular network: Deployment solutions, sensitivity and tradeoffs," in *Proc. WiAd,* 2011, pp. 42-47.

[89] A.A. Abdulkafi, T.S. Kiong, D. Chieng, A. Ting, J. Koh, "Energy efficiency improvements in heterogeneous network through traffic load balancing and sleep mode mechanisms," *Wireless Pers. Commun.,* vol. 75, no. 4, pp. 2151-2164, Apr. 2014.

[90] J. Wu, S. Jin, L. Jiang and G. Wang, "Dynamic switching off algorithms for pico base stations in heterogeneous cellular networks," *EURASIP J. Wireless Commun. Netw.,* vol. 2015, no. 1, pp. 1-18, Dec. 2015.

[91] H. Narasimhan, R. Kashyap, A.V. Vikram, R.C. Deepthi and P. Wali, "Greedy Search Algorithm Based Two Stage Energy Saving Technique for Dense LTE-A HetNets," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, India, 13-16, Sept. 2017.

[92] N. Obaid and A. Czylwik, "Energy efficiency analysis of dense pico cell deployments," in *Proc. 18th Int. OFDM Workshop (InOW),* Aug. 2014, pp. 1-6.

[93] B. Suman and P. Kumar, "A survey of simulated annealing as a tool for single and multiobjective optimization," *J. Oper. Res. Soc.,* vol. 57, no. 10, pp. 1143-1160, 2006.

[94] F. Busetti, "Simulated Annealing Overview," Rep. 2003.

[95] C. Liu, B. Natarajan and H. Xia, “Small cell base station sleep strategies for energy efficiency," *IEEE Commun. Veh. Technol.,* vol. 65, no. 3, pp. 1652-1661, Mar. 2016.

[96] O. Blume, H. Eckhardt, S. Klein, E. Kuehn, W.M. Wajda, “Energy savings in mobile networks based on adaptation to traffic statistics," *Bell Labs Tech. J.* 15(2), 77-94, 2010.

[97] H. Kwon and T. Birdsall, “Channel capacity in bits per joule," *IEEE J. Ocean. Eng.,* vol. 11, no. 1, pp. 97-99, Jan. 1986.

[98] Silva JF, Serra D. “A capacitated facility location problem with constrained backlogging probabilities," *International Journal of Production Research, Forthcoming*, 2007; 45(21): 5117-5134.

[99] Ling-Yun Wu, Xiang-Sun Zhang, and Ju-Liang Zhang, “Capacitated facility location problem with general setup cost," *Computers & Operations Research,* 33(5):1226-1241, 2006.

[100] Hindi, K. S., and K Pienkosz. “Efficient Solution of Large Scale, Single-Source, Capacitated Plant Location Problems," *The Journal of the Operational Research Society,* vol. 50, no. 3, 1999, pp. 268-274. JSTOR, www.jstor.org/stable/3010690. Accessed 22 Jan. 2020.

[101] S. Brooks and B. Morgan, “Optimization using simulated annealing,” *Journal of the Royal Statistical Society. Series D (The Statistician),* vol. 44, no. 2, pp. 241-257, 1995.

[102] D. Komnakos, A. Rouskas, A. Gotsis, “Energy Efficient Base Station Placement and Operation in Mobile Networks,” in *the Proc of the 19th European Wireless Conference (EW 2013),* Guildford, United Kingdom, April 2013.

[103] O. Blume, A. Ambrosy, M. Wilhelm, and U. Barth, “Energy efficiency of LTE networks under traffic loads of 2020,” in *Proc. of the 10th Int. Symp. on Wireless Communication Systems,* Aug. 2013, pp. 1-5.

[104] S. Sun, C. Zhang, and Y. Guoqiang, “A Bayesian network approach to traffic flow forecasting,” *IEEE Intell. Transp. Syst. Mag.,* vol. 7, no. 1, pp. 124 -132, Mar. 2006.

[105] K. Takeshita, M. Yokota, and K. Nishimatsu, “Early network failure detection system by analyzing Twitter data," in *Proc. IEEE Int. Symp. Integr. Netw. Manag.* (IM), Ottawa, ON, Canada, 2015, pp. 279-286.

[106] J. Pérez-Romero, J. Sánchez-González, O. Sallent, and R. Agustí, “On Learning and Exploiting Time Domain Traffic Patterns in Cellular Radio Access Networks,” in *Machine Learning and Data Mining in Pattern Recognition,* pp. 501-515, Springer, 2016.

[107] F. Somaa, C. Adjih, I. Korbi, L. Saidane, “A Bayesian model for mobility prediction in wireless sensor networks,” in *5th IFIP International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN 2016),* 2016.

[108] T.T. Son, H.L. Minh, G. Sexton, N. Aslam and Z. Ghassemlooy, “Bayesian Model for Mobility Prediction to Support Routing in Mobile Ad-Hoc Networks,” in *Proc of the 24th IEEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC’ 13,* UK, Sept. 2013.

[109] F. Xu, Y. Li, H. Wang, P. Zhang, and D. Jin, “Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment,” *IEEE/ACM Trans. Netw.,* pp. 1-15, 2016.

[110] P.V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, “A Survey of Machine Learning Techniques Applied to Self Organizing Cellular Networks,” *IEEE Commun. Surv. Tutorials,* no. c, pp. 1-1, 2017.

[111] Shadab Adam Pattekari and Alma Parveen, “Prediction system for heart disease using Naive Bayes”, *International Journal of Advanced Computer and Mathematical Sciences*, vol.3, pp 290-294, 2012.

[112] S. D. Jadhav and H. P. Channe, “Comparative Study of K-NN , Naive Bayes and Decision Tree Classification Techniques”, *Int. J. Sci. Res.,* vol. 5, no. 1, pp. 1842-1845, 2016.

[113] T.M. Mitchell, “Bayesian Learning," in Machine Learning, McGrawHiII, 1997, pp. 154 - 200.

[114] Z. Niu, S. Zhou, Y. Hua, Q. Zhang, and D. Cao, “Energy-Aware Network Planning for Wireless Cellular System with Inter-Cell Cooperation,” *IEEE Trans. Wirel. Comm.,* vol. 11, no. 4, pp. 1412-1423, Apr. 2012.

[115] K. Okino, T. Nakayama, C. Yamazaki, H. Sato, and Y. Kusano, “Pico cell range expansion with interference mitigation toward LTE-Advanced heterogeneous networks,” in *Proc. IEEE Int. Workshop on Heterogeneous Networks,* June 2011.

[116] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, “An overview of load balancing in HetNets: Old myths and open problems,” *IEEE Wireless Commun.,* vol. 21, no. 2, pp. 18-25, Apr. 2014.

[117] Nokia Siemens Networks, Nokia, “Aspects of Pico Node Range Extension,” 3GPP TSG RAN WG1 Meeting 61, R1- 103824, 2010, available: http://goo.gl/XDKXI.

[118] A. Damnjanovic et al., “A Survey on 3GPP Heterogeneous Networks,” *IEEE Wireless Commun.,* vol. 18, no. 3, June 2012, pp. 10-21.

[119] Q. Ye et al., “User Association for Load Balancing in Heterogeneous Cellular Networks,” *IEEE Trans. Wireless Commun.,* vol. 12, no. 6, June 2013, pp. 2706-16.

[120] P.T. V. Bhuvaneswari, S. Indu, N. Lathiffa Shifana, et al., “An analysis on cell range expansion in 4G LTE networks,” in *2015 3rd international conference on signal processing, communication and networking (ICSCN 2015),* 2015.

[121] T. Qu, D. Xiao, D. Yang, W. Jin, Y. He, "Cell selection analysis in outdoor heterogeneous networks," *3rd International Conference on Advanced Computer Theory and Engineering,* 2010, vol 5, pp. no.554-557.

[122] F.H. Tseng, L.D. Chou, H.C. Chao, W.J. Yu, "Set cover problem of coverage planning in LTE-Advanced relay networks," *International Journal of Electronic Commerce Studies,* Vol.5, No.2, pp.181-198, 2014.

[123] A. Lobinger, S. Stefanski, T. Jansen, and I. Balan, "Load balancing in downlink lte self-optimizing networks," in *VTC Spring,* 2010, pp. 1-5.

[124] M. Shirakabe, A. Morimoto, and N. Miki, "Performance evaluation of inter-cell interference coordination and cell range expansion in heterogeneous networks for LTE-advanced downlink," in *ISWCS,* pp. 844-848, Nov. 2011.

[125] S. Mertens, "The Easiest Hard Problem: Number Partitioning," in *Computational Complexity and Statistical Physics*, A. Percus, G. Istrate, and C. Moore, Eds. Oxford University Press, 2006, pp. 125?139.

[126] R.E. Korf, "Multi-way number partitioning," in *Proceedings of the 21st international joint conference on rtifical intelligence*. Morgan Kaufmann Publishers Inc., 2009, pp. 538-543.

[127] Chakravarty, A.K., Orlin, J.B., and Rothblum, U.G., "A partitioning problem with additive objective with an application to optimal inventory groupings for joint replenishment," *Operations Research 30* (1982) 1018-1022.

[128] M. Eguizabal and A. Hernandez, "Interference management and cell range expansion analysis for LTE picocell deployments," in *Proc. PIMRC,* pp. 1592-1597, Sept 2013.

[129] H. Zhang, X. Qiu, L. Meng, and X. Zhang, "Design of distributed and autonomic load balancing for self-organization LTE," in *Proc. IEEE 72nd VTC,* 2010, pp. 1?5.