Compression and Recovery in Cell-free Cloud Radio Access Network

Ming Zhao

Master of Science (by Research)

University of York

Electronic Engineering

September 2018

Abstract

Cloud radio access network (C-RAN) is an evolved network architecture for future mobile communication systems. It aims to provide higher spectral efficiency, lower energy consumption and reduced cost of operations and maintenance for the network, which will enable the operators to not only satisfy growing user demands, but provide new services and applications. However, the huge load on the fronthaul network which connects the baseband unit (BBU) and a large number of remote radio heads (RRHs) is a significant challenge. To improve the fronthaul performance, a data compression and recovery scheme based on compressive sensing is proposed in this thesis. First, the theory of compressive sensing is studied, including the essential principles, standard compressive sensing model, potential measurement matrices, etc. Several popular recovery algorithms in compressive sensing are demonstrated in detail. Secondly, a compression and recovery scheme is proposed for the uplink of a cell-free C-RAN system. In the proposed scheme, compressive sensing is applied by exploiting the sparsity of user data. In particular, the multi-access fading in this system is incorporated into the formulation of the compressive sensing model. The aggregated measurement matrix which contains both the channel matrix and the fronthaul compression matrix is shown to satisfy the restricted isometry property (RIP) condition. Furthermore, two different recovery algorithms, basis pursuit denoising (BPDN) and sparsity adaptive matching pursuit (SAMP), are used respectively for estimating the sparse signals. The major advantage is that they do not require the sparsity of user data as a prior information during the process of signal recovery. It allows easy applications in many practical scenarios where the number of non-zero elements of the signals is not available. The simulation results show that the proposed scheme can efficiently alleviate the heavy burden on the fronthaul network, and meanwhile provide stable signal recovery for this system.

Contents

Abstract **i**

List of Figures **vi**

List of Tables **viii**

Acknowledgements **ix**

Declaration **x**

1 Introduction **1**

1.1 Overview 1

1.2 Contributions 3

1.3 Thesis Outline 3

1.4 Notations 4

2 Literature Review **5**

2.1 Introduction 5

2.1.1 Background 6

2.2 Architecture of C-RAN 7

2.2.1 Evolution of BS Architecture 7

2.2.2 C-RAN Architectures Proposed by Industries 9

2.2.3 C-RAN Architectures Proposed by Academic Researchers 11

2.3 Advantages of C-RAN 12

2.4 Technical Challenges and Open Issues of C-RAN 13

2.5 Key Technologies of C-RAN 14

2.5.1 Compression in the Uplink 15

2.5.2 Compression in the Downlink 18

3 Theory of Compressive Sensing **21**

3.1 Introduction 22

3.2 Basic Concepts 24

3.2.1 Standard Compressive Sensing Model 24

3.2.2 Fundamental Principles 24

3.3 Robustness of Compressive Sensing 26

3.3.1 Restricted Isometry Property 27

3.3.2 Potential Measurement Matrices 28

3.4 Recovery Algorithms 29

3.4.1 Basis Pursuit Denoising 29

3.4.2 Greedy Algorithms 30

3.4.3 Comparison Between BPDN and Greedy Algorithms 36

4 Cell-free C-RAN System with Compressive Sensing **38**

4.1 Introduction 38

4.2 Uplink C-RAN System Model 39

4.3 Proposed Compression and Recovery Scheme 41

4.3.1 Fronthaul Compression and Forward 42

4.3.2 Signal Recovery at BBU Pool 42

4.3.3 Overall Compression and Recovery Scheme 47

4.4 System Analysis 48

4.5 Numerical Results and Discussion 49

4.5.1 Parameter Setup 49

4.5.2 The Percentage of Exact Recovered Signals Versus Number of Measurements 50

4.5.3 The Percentage of Exact Recovered Signals Versus Sparsity . . . 54

4.6 Summary 55

5 Conclusions and Further Work **57**

5.1 Conclusions 57

5.2 Further Work 58

Glossary **60**

Bibliography **62**

List of Figures

2.1 Forecast of global data traffic between 2016 and 2021 [2] 6

2.2 BS architecture evolution [4]. 8

2.3 C-RAN architecture proposed by China Mobile Research Institute. 10

4.1 Uplink cell-free C-RAN system 40

4.2 The flow chart of signal recovery using BPDN 43

4.3 The overall compression and recovery scheme 47

4.4 Percentage of recovered signals versus number of measurements under number of active users and the number of active users is $s = 20$. (a) $M = 20$. (b) $M = 50$. 52

4.5 Percentage of recovered signals versus number of measurements at each RRH via BPDN, for various numbers of RRHs $M = 20, 30, 40, 50, 60$ and $s = 20$ active users. 53

4.6 Percentage of recovered signals versus number of measurements at each RRH via SAMP (with step size $t = 5$), for various numbers of RRHs $M = 20, 30, 40, 50, 60$ and $s = 20$ active users. 54

4.7 Percentage of recovered signals versus number of active users. Here, the number of RRHs $M = 50$ and the number of measurements at each RRH $R = 4$. 55

List of Tables

2.1 Summary of literature related to uplink fronthaul compression 16

2.2 Summary of literature related to downlink fronthaul compression 18

3.1 Comparison of complexity of different recovery algorithms 37

Acknowledgements

I would firstly like to show my sincere gratitude to my supervisor, Prof. Alister G. Burr, for his helpful support, supervision and guidance on my research. I have learned a lot from his preciseness and immense knowledge in communications.

I would also like to acknowledge my thesis advisor, Dr. Kanapathippillai Cumanan, whose insightful advice have benefited me.

I am also grateful to my fellow research students, for their helpful comments, discussions and of course friendship.

Finally, my profound gratitude goes to my parents for their unconditional support, endless love and continuous encouragement throughout my research and my life in general. Thank you.

Declaration

This work has not previously been presented for an award at this, or any other, University.

All work presented in this thesis as original is so, to the best knowledge of the author. References and acknowledgements to other researchers have been given as appropriate.

Chapter 1

Introduction

Contents

1.1 Overview 1

1.2 Contributions 3

1.3 Thesis Outline 3

1.4 Notations 4

1.1 Overview

The demand for higher data rate and higher quality of service, resulting in greater costs to build, operate and upgrade the traditional radio access network (RAN), has caused operators to seek alternative approaches to access network architecture. Cloud radio access network (C-RAN) is such an evolved network architecture that can address these challenges for operators, as well as satisfy growing user demands [1].

A typical C-RAN architecture is comprised of three parts: a pool of baseband units (BBUs) where the baseband processing is centralised, a large number of remote radio heads (RRHs) with co-located antennas which are distributed at the remote sites, and the high bandwidth, low latency optical transport network connecting the BBU pool and RRHs. This architecture provides higher spectral efficiency and lower power consumption

thanks to resource sharing in the BBU pool via base station virtualisation [1]. It also allows centralised management and hence reduces the cost of operation and maintenance substantially.

However, the transport network between the BBU pool and the RRHs, which is also called the fronthaul link, brings a challenging problem for practical implementation: the huge load on the fronthaul due to a huge amount of I/Q data that needs to be transmitted in real time. This problem has attracted a great deal of research interest. Different strategies and techniques have been developed for dealing with this issue, including distributed source coding, joint compression, compressive sensing based compression strategy, etc [14]. In this thesis, we will focus primarily on a compressive sensing based compression strategy.

The theory of compressive sensing is first studied. Based on a standard compressive sensing model, the key notions of sparsity and incoherence are introduced respectively. A more practical model that takes data corruption into account is further discussed prior to the introduction of the restricted isometry property (RIP), which is a powerful tool to study the robustness of compressive sensing. Furthermore, two different types of popular recovery algorithms in compressive sensing, namely, basis pursuit denoising (BPDN) [40] and greedy algorithms are illustrated in detail.

Based on the investigation of compressive sensing, a compression and recovery scheme is proposed for an uplink C-RAN scenario in order to relieve the heavy burden on the fronthaul link. As the uplink signals tend to be sparse in some particular scenarios such as massive machine-type communication (mMTC) system [54], compressive sensing can be easily applied to process the user data at each RRH and then reconstruct the original data at the BBU pool. More precisely, the proposed scheme is composed of three stages: 1) signal gathering at each RRH, 2) fronthaul compression, and 3) signal recovery at the BBU pool. In particular, in the signal recovery stage, BPDN and another greedy algorithm called sparsity adaptive matching pursuit (SAMP) [51] are used respectively for estimating the sparse signals. The proposed scheme can fulfil robust signal recovery as well as bring down the fronthaul load effectively.

1.2 Contributions

The major contribution in this thesis is summarised as follows:

- Exploiting the sparsity of user data, a compressive sensing based compression and recovery scheme is developed for the uplink of cell-free C-RAN system. It has been shown in the numerical results that the proposed scheme can achieve robust recovery of the sparse signals and effective fronthaul load reduction.

1.3 Thesis Outline

The content of the thesis is organised as follows:

- In Chapter 2, a literature review of C-RAN is presented which includes the background on traditional RAN, different architectures of C-RAN proposed by industries and researchers, advantages of C-RAN, technical challenges and open issues of C-RAN, as well as advanced technologies in C-RAN corresponding to the existing challenges, in which we mainly focus on the state of the art techniques in fronthaul compression.
- In Chapter 3, a sampling approach called compressive sensing is introduced in detail. An overview of compressive sensing and its applications is given first. Secondly, the standard compressive sensing model is given along with several essential notions in compressive sensing. Thirdly, the robustness of compressive sensing is studied. Finally, two different kinds of recovery algorithms are demonstrated, including the convex optimisation method (i.e., BPDN) and the greedy algorithms, and a comparison between these two kinds of approaches is given.
- In Chapter 4, we consider an uplink C-RAN system where compressive sensing is applied to improve the fronthaul performance by leveraging the sparsity of user data. A compression and recovery scheme is proposed for this system. In particular, we adopt the BPDN and SAMP algorithms to recover the sparse uplink signals. Lastly, the system analysis and numerical results of proposed scheme are presented.

- In Chapter 5, conclusions and the possible future work based on this thesis are presented.

1.4 Notations

In this thesis, we use uppercase boldface letters, lower boldface letters and plain letters to denote matrix, vector and scalar variables respectively. The boldface **0** represents either a zero vector or a zero matrix. The operators $|\cdot|$, $(\cdot)^T$, $(\cdot)^H$, $\#|\cdot|$, $(\cdot)^{-1}$, $(\cdot)^\dagger$ represent the modulus, transpose, conjugate transpose, cardinality, inverse, Moore-Penrose pseudoinverse respectively. $\|\cdot\|_1$, $\|\cdot\|_2$ denote the l_1 norm and l_2 norm respectively. $E(\cdot)$ represents the expectation value.

Chapter 2

Literature Review

Contents

2.1 Introduction 5

2.2 Architecture of C-RAN 7

2.3 Advantages of C-RAN 12

2.4 Technical Challenges and Open Issues of C-RAN 13

2.5 Key Technologies of C-RAN 14

2.1 Introduction

The literature review of cloud radio access network (C-RAN) system development is presented in this chapter. First, the background on the traditional radio access network (RAN) is summarised. Following this, the evolution of the base station (BS) and different architectures of C-RAN are introduced alongside the description of the advantages of C-RAN. Finally, the existing technological challenges, open issues and corresponding state-of-the-art techniques in C-RAN are illustrated respectively.

2.1.1 Background

Global mobile data traffic has been growing significantly over the past several years. According to the white paper from Cisco [2], the mobile data traffic has risen 18-fold between 2011 and 2016. And it is predicted that the global data traffic will increase at a compound annual growth rate (CAGR) of 47 percent from 2016 to 2021, reaching 49 exabytes per month by 2021, as shown in Fig. 2.1 [2]. This strong growth of mobile data traffic results in serious challenges to the traditional RAN.

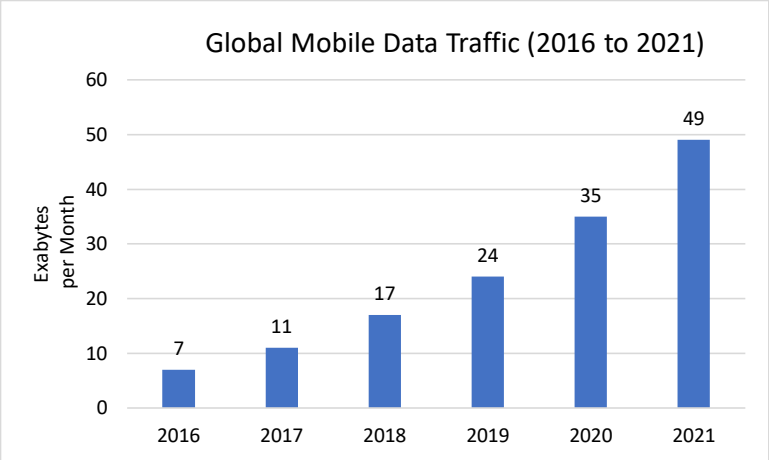


Figure 2.1: Forecast of global data traffic between 2016 and 2021 [2]

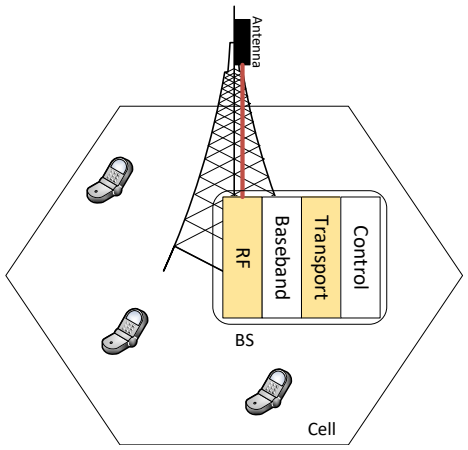
In the traditional RAN, only a fixed number of mobile terminals are served by each BS, which means each isolated BS only processes signals within a small service area. As a result, operators have to increase the number of BSs to offer larger network coverage and higher capacity, which gives rise to huge power consumption, and thereby dramatically increasing the capital expenditure (CAPEX) and the operating expenditure (OPEX) which mainly refer to the cost of network construction, operations as well as maintenance [1]. On the other hand, operators are facing a worrying situation that the revenue is not growing at the same pace as the increasing cost of network construction. Besides, interference between adjacent BSs which share the same frequency resource makes it difficult to achieve interactive communications so as to improve spectrum utilisation rate. Therefore, mobile operators must find an appropriate network architecture with advanced techniques to cope with the requirements of high spectral efficiency and low power consumption as well as to keep profitable growth.

2.2 Architecture of C-RAN

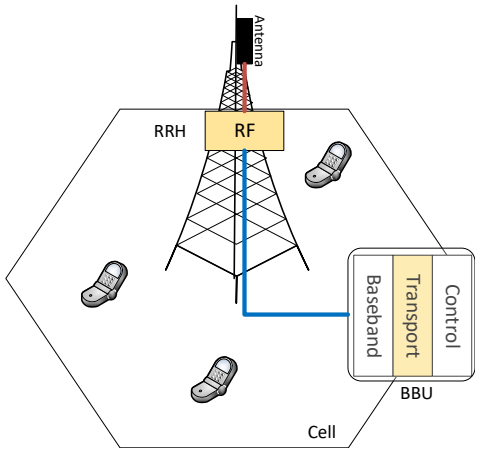
C-RAN is recognised as a promising solution to address the challenges mentioned above. The concept of C-RAN was first proposed by IBM, called wireless network cloud (WNC) in [3], and described in detail in [1]. To start with, the evolution of BS is illustrated, prior to a description of general C-RAN architecture and a survey on several specific C-RAN architectures proposed by either industries or researchers.

2.2.1 Evolution of BS Architecture

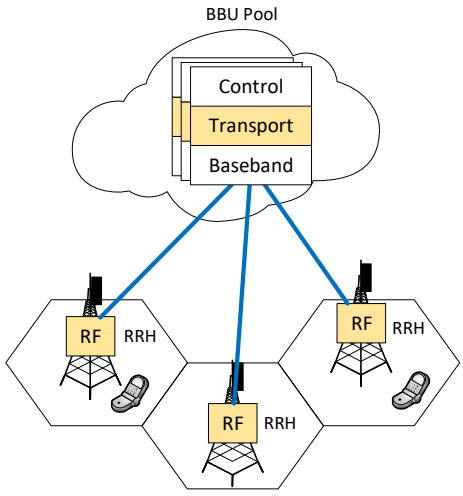
The BS architecture evolution is shown in Fig. 2.2 [4]. In traditional BS architecture, as shown in Fig. 2.2a, both radio and baseband processing functions are integrated into one BS. This is the popular network deployment in 1G and 2G networks. In 3G and 4G cellular network, the BS is split into a signal processing unit and a radio unit, as shown in Fig. 2.2b. The signal processing unit is called the baseband unit (BBU), and the radio unit is described as a remote radio head (RRH) which accomplishes radio frequency (RF) processing, digital-to-analog conversion, analog-to-digital conversion, power amplification and filtering [5]. In such an architecture, the BBU can be deployed at a distance of up to 40 km from the RRH. It means the BBU can be placed in an easily accessible site whereby the cost of site rental and maintenance can be saved. In 4G and beyond, BBUs may be aggregated into a BBU pool to optimise the BBU utilisation and achieve resource sharing among a large number of virtual BSs, as shown in Fig. 2.2c. This yields the general architecture of C-RAN, which is mainly composed of three parts: the RRHs with co-located antennas deployed at a remote site; the BBU pool that consists of virtualisation technology and high-performance processors to perform baseband processing; and the high bandwidth low-latency fronthaul link connecting the BBU pool and the RRHs.



(a) Traditional BS architecture



(b) BS with RRH



(c) C-RAN with RRH

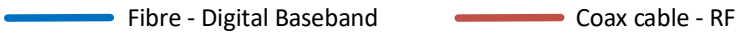


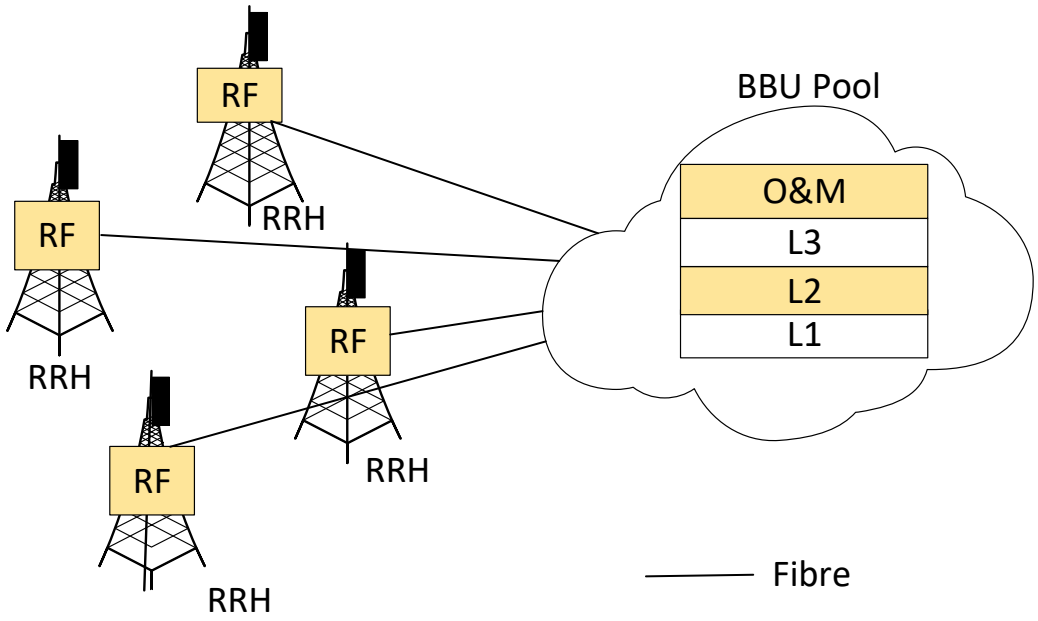
Figure 2.2: BS architecture evolution [4].

In addition to this general design, the operators, equipment vendors as well as researchers are dedicated to optimising C-RAN architectures.

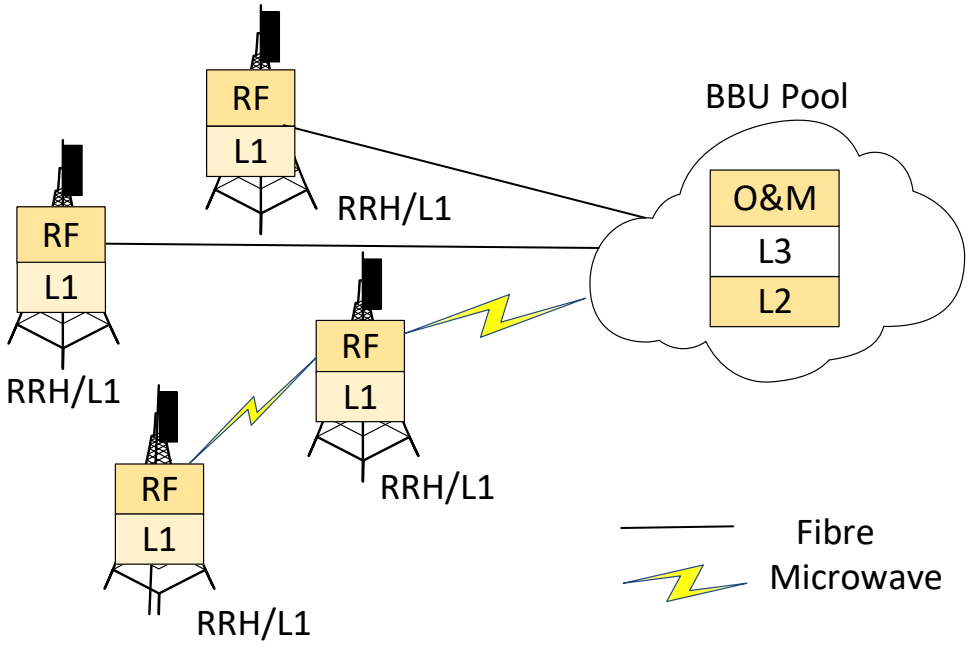
2.2.2 C-RAN Architectures Proposed by Industries

After being firstly proposed by IBM under the name of WNC [3] and further developed by China Mobile Research Institute [1], different enhancements to the C-RAN architecture have been proposed by other operators and equipment vendors [6–9]. To save fibre resources and cope with C-RAN bearer requirements, ZTE provided three different solutions, namely, enhanced fibre connection, colored fibre connection and optical transport network bearer for different scenarios [6]. The enhanced fibre connection solution has been successfully commercialised, which benefits the widespread C-RAN deployment. Targetted at scalable, larger-scale BBU pools, a GPP based C-RAN architecture is proposed by Intel [7], which can use computation resources and switch bandwidth as needed and thus lower the overall power consumption compared to traditional BS. In [8], Alcatel-Lucent introduced the novel concept of the cloud BS exploiting the elements of cloud computing, which aims at reducing the required processing resources without sacrificing system performance. In the same year, another concept, called RAN-as-a-Service (RANaaS), was presented by Telecom Italia [9]. This network architecture enables more flexibly centralised RAN functionalities through a cloud-type platform and hence better optimises the resources as well as managing the network.

The key difference among these architectures presented by the industries lies in the different function splitting between the BBU pool and the RRHs. For instance, China Mobile Research Institute classifies C-RAN solutions into two types: ‘fully centralised’ C-RAN architecture and ‘partially centralised’ C-RAN architecture [1]. In ‘fully centralised’ C-RAN architecture, as depicted in Fig. 2.3a, physical layer (PHY/L1), medium access control layer (MAC/L2) and network layer (L3) functions are all located in the BBU pool. It has the advantages of more flexible network upgrading, easier capacity expansion, more efficient multi-cell collaborative processing and maximising resource sharing. It can also reduce the cost of installing RRHs as there are few digital processing devices required at RRHs. However, the high bandwidth between BBU and RRHs for transmitting in-phase and quadrature (I/Q) signals becomes a heavy burden.



(a) Fully centralised C-RAN architecture



(b) Partially centralised C-RAN architecture

Figure 2.3: C-RAN architecture proposed by China Mobile Research Institute.

On the other hand, in the 'partially centralised' C-RAN architecture, as shown in Fig. 2.3b, the baseband functions are moved to the RRHs, while all other higher layer functionalities still reside in the BBU pool. This solution can effectively relieve the burden on fronthaul links, as the demodulated signal only occupies $1/20$ – $1/50$ of the bandwidth

required for the modulated one [1]. But it is less convenient for collaborative processing because functions in PHY are integrated into the RRHs.

2.2.3 C-RAN Architectures Proposed by Academic Researchers

In addition to the industry, the academic community has also proposed advances in C-RAN architecture. A three-layer logical structure of C-RAN, including a physical plane, a control plane and a service plane, is proposed in [10]. With the emphasis on the concept of the service cloud, the proposed structure is designed to achieve more efficient centralised processing. Based on radio-over-fibre (RoF) technology, a multi-service small cell C-RAN architecture is proposed in [11], which can efficiently simplify the design of RRH by transmitting analogue RF signals over the fibre-optic link. Besides, a more flexible configuration of the fronthaul network can be achieved by integrating it with wavelength division multiplexing (WDM) technology. In [12], a heterogeneous cloud radio access network (H-CRAN) is proposed as a cost-efficient paradigm to improve both spectral and energy efficiencies by incorporating cloud computing technology into heterogeneous networks (HetNets). The authors not only demonstrate key H-CRAN techniques, such as cloud-computing-based coordinated multipoint and large-scale cooperative multiple antenna processing, but also discuss the potential challenges and open issues in H-CRAN. To deal with the existing problems of H-CRAN, a fog-computing-based RAN (F-RAN) architecture is further proposed in [13]. By making full use of collaborative radio signal processing and cooperative radio resource management in edge devices, F-RAN is envisioned to mitigate the burden on the fronthaul.

In a nutshell, all the architectures mentioned above can be alternative deployment solutions in different scenarios according to the situation of the networks. The next generation (5G/6G) mobile communications are appealing for more enhancements in existing C-RAN architectures to meet the increasing requirements. In this thesis, the study will mainly focus on the ‘fully centralised’ C-RAN architecture.

2.3 Advantages of C-RAN

With above survey on the architecture of C-RAN, the benefits of C-RAN are summarised as follows [1, 4]:

- Cost-saving on CAPEX and OPEX
  Thanks to the gathering of BBUs and other computation equipment within a few big rooms, centralised management can be easily achieved and the cost of operation and maintenance can be substantially reduced. Secondly, the functionality of the RRHs in a fully centralised C-RAN architecture is simpler. They can be equipped with less site support and their size are smaller. This means operators can accelerate the installation of RRHs, and consequently enables cost saving on the network construction as well as site rental.

- Eco-friendly Infrastructure
  C-RAN is regarded as a green infrastructure [1]. Firstly, energy on power consumption and air conditioning can be significantly decreased as fewer BBUs are required in a C-RAN system compared to a traditional RAN. Moreover the air conditioning of radio module located in RRHs can be reduced because RRHs are naturally cooled by air [4]. Secondly, in the BBU pool, resources can be shared among a great number of virtual BSs, which maximises the utilisation rate of processing data and thus brings down the power consumption. Lastly, when some BBUs are in lower traffic period (e.g., at night), they can be selectively turned to low power mode or even switched off without affecting the overall service.

- Spectrum Utilisation Improvement
  With the centralised processing in C-RAN system, virtual BSs in a BBU pool can smoothly share the traffic data, channel estimation information (CSI) of each active user and control information of users. Furthermore, cooperative processing schemes, such as cooperative multi-point processing (CoMP) and enhanced inter-cell interference coordination (eICIC), can be easily adopted to mitigate the inter-cell interference (ICI), thereby improving the spectral efficiency and throughput.

- Adaptability to Non-uniform Traffic and Scalability
  The data traffic varies during a day, and the peaks of traffic change location at

different hours as users move around. C-RAN is able to cope with the non-uniform traffic thanks to the load balancing feature in the BBU pool. RRHs in the same BBU pool can be selected dynamically to offer services according to the movement of users, and thus increase the overall utilisation rate of the BBU pool, providing a better user experience. Furthermore, with any of the above C-RAN architectures, operators can easily and quickly deploy and upgrade their networks. Operators only need to add some new RRHs to the already existing BBU pool to offer more service areas. The network capacity can then be improved by either splitting the existing cell or adding additional RRHs.

2.4 Technical Challenges and Open Issues of C-RAN

Though C-RAN brings a lot of benefits in respect of cost, energy efficiency and spectral efficiency, it also has a number of technical challenges that need to be solved before commercial deployment [1, 4, 15].

- High Bandwidth, Low Latency and Low Cost Fronthaul Transport Network
  In the fully centralised C-RAN architecture, the optical links between RRHs and BBU pool have to support a high bandwidth as a huge amount of I/Q data is transmitted in real time, which will lead to large overheads. Besides, the transport network also needs to meet the low latency and strict jitter requirements. For example, as the most advanced CoMP scheme, joint transmission (JT) requires only 0.5 μs timing accuracy in base station synchronisation, which is a very tight constraint [4].

- Advanced Coordinated Processing and BBU Pool Interconnection
  To improve the spectrum efficiency and mitigate the interference, advanced cooperative processing algorithms should be executed among virtual BSs with the support of sharing the user data, channel information and scheduling information of virtual BSs. Besides, radio resource sharing algorithms incorporating optimal clustering algorithms should be developed to handle the interference and also reduce the scheduling complexity.

  Security is crucial in a scenario in which many BBUs are centralised in one entity.

To achieve high reliability and flexible resource allocation, the network enabling interconnection of BBUs must be designed to have a reliable, flexible and extensible topology. Data from any RRH can be processed by any available BBU through this network, and therefore the overall performance of the system will not be affected by any individual BBU failure.

- Virtualisation for BBU Pool
  Virtualisation plays a vital role in the implementation of C-RAN. It enables the processing data to be flexibly and dynamically allocated among virtual BS entities. The virtualisation scheme is expected to realise real-time processing and dynamic capacity allocation in order to cope with dynamically changing cell load in C-RAN system.

- Other Open Issues
  In addition to the above mentioned challenging works, there are still many issues need to be solved, including edge caching, physical layer security, field trials, and so on [14, 15].

2.5 Key Technologies of C-RAN

The state-of-the-art technologies in C-RAN corresponding to the above challenges are categorised as fronthaul compression, large scale collaborative processing, radio resource allocation optimisation, BBU interconnection as well as virtualisation technology [1, 4, 14, 15]. Since this thesis mainly focusses on the fronthaul compression in C-RAN, this section will introduce recent advanced fronthaul compression techniques only.

In general, current popular fronthaul compression technologies can be sorted into quantisation based compression, compressive sensing based compression and hybrids of independent compression and other techniques. To be specific, a quantisation based compression strategy is executed by quantising the original data and then optimising the quantisation noise level to improve the fronthaul performance. This strategy is mainly composed of point-to-point compression, distributed source coding and joint compression, in which distributed source coding and point-to-point compression are usually utilised in the

uplink of C-RAN, while joint compression and point-to-point compression are used in the downlink of C-RAN [14]. For the compressive sensing based compression strategy, the compression is realised via a linear operation at each RRH, i.e., multiplying the original data with a lower-dimensional matrix. compressive sensing based compression is usually used in the uplink of C-RAN.

2.5.1 Compression in the Uplink

In the uplink, the received signals at each RRH are compressed before being transmitted to the BBU pool through the fronthaul link. As mentioned above, there are three main ways to complete the compression: point-to-point compression, distributed source coding and compressive sensing based compression. Distributed source coding exploits the correlation among signals received by neighbouring RRHs to reduce the fronthaul transmission rate, but at the expense of high complexity [15]. To realise the compression with low complexity, point-to-point compression will be an alternative. In addition, compressive sensing based compression can be used in the scenario where the received signal is sparse or compressible. The uplink frontahul compression approaches are briefly summarised in Table 2.1.

Table 2.1: Summary of literature related to uplink fronthaul compression

Compression approach	Literature	System model	Main contribution
Point-to-point compression and distributed compression	[16]	Multiple multi-antenna RRHs and multi-antenna user model with limited per RRH fronthaul capacity	Review the performance of conventional point-to-point compression and provide a comparison between point-to-point compression and distributed compression
	[18]	Multiple single-antenna RRHs and single-antenna user model with a limited sum fronthaul capacity	Perform efficient optimisation of the quantisation noise levels under high SQNR
	[19]	Multiple multi-antenna RRHs and multi-antenna user model with limited per RRH fronthaul capacity	Propose a joint design of fronthaul compression and beamforming
Distributed source coding	[17]	Multiple multi-antenna RRHs and multi-antenna user model with limited per RRH fronthaul capacity	Propose a robust compression scheme that can make up for the performance loss caused by imperfect statistical information
Compressive sensing based compression	[20]	Multiple single-antenna RRHs and single-antenna user model without fronthaul capacity constraint	Exploit the sparsity of uplink user data and propose a distributed fronthaul compression scheme and joint signal recovery algorithm
	[21]	Multiple single-antenna RRHs and single-antenna user model with limited per-RRH fronthaul capacity	Explore the correlation of received signals and propose a compression scheme for multi-cluster C-RAN

Point-to-point compression [16] is a conventional fronthaul compression method which quantises signals for different RRHs separately. The basic concepts and key information-theoretic results are presented in [16] along with the comparison with other advanced fronthaul compression schemes. It is shown that though standard point-to-point compression reaps the benefit of separate processing, it ignores the correlation among received signals at each RRH, and thus fails to achieve the optimal performance. Therefore, distributed compression that adopts joint signal processing is expected to have an advantage over point-to-point compression.

A lot of literature has presented the work of compression in uplink C-RAN by utilising distributed source coding. In [17], a robust compression scheme is proposed to address the problem that the errors in statistical information about the correlation of received signals at different RRHs will affect the performance of distributed source coding. Besides, a joint RRH selection and compression method is further developed in order to improve the energy efficiency of the system. The simulation results verify that the proposed robust

compression scheme can effectively make up for the performance loss caused by imperfect statistical information. In [18], a quantisation noise level optimisation problem is formulated to maximise weighted sum rate under a limited sum fronthaul capacity by adopting either distributed Wyner-Ziv coding or single-user compression (i.e., point-to-point compression). The proposed algorithm can jointly optimise the quantisation noise levels of all RRHs, which is different from [17] where the optimisation problem of quantisation noise level is solved at each RRH. It is observed that setting the quantisation noise level to be proportional to the background noise level is near optimal for maximising the sum rate in the high signal-to-quantisation-noise-ratio (SQNR) regime. With a proper choice of quantisation noise levels, the proposed Wyner-Ziv coding based scheme can achieve the sum capacity within a constant gap. A similar result can be obtained for single-user compression under a diagonally dominant channel condition. However, [18] only considers the uplink single-input single-output (SISO) C-RAN model. To extend to the multiple-input multiple-output (MIMO) case, [19] considers a uplink C-RAN model where both the RRHs and users are equipped with multiple antennas. A joint optimisation problem of the transmit beamformer and quantisation noise covariance matrix for weighted sum rate maximisation is formulated. An effective algorithm is proposed for tackling this joint optimisation problem via Wyner-Ziv coding and single-user compression. Moreover, a separate design of beamforming and fronthaul compression is developed to find the near optimal maximum weighted sum rate at high SQNR, which is shown to perform well in practical multicell networks.

Instead of reducing the fronthaul loading through quantisation, the compressive sensing based compression approach exploits the sparsity of the received signals in uplink C-RAN to fulfil this goal. In [20], a fronthaul compression strategy at each RRH and an end-to-end signal recovery algorithm at BBU pool are proposed, exploiting the sparsity of user data. The authors analyse the performance of the proposed signal recovery algorithm and remark that the aggregate measurement matrix which consists of the channel matrix and distributed compression matrix can still satisfy the restricted isometry property (RIP), which is one of the important characteristics of compressive sensing. Moreover, the tradeoff between fronthaul loading and sum capacity in a C-RAN system is derived. Unlike [20] where the user data is assumed to be sparse intrinsically, the authors in [21] consider the scenario where the user data within one cluster is non-sparse but statistically correlated, and thus the gathered signal vector can be sparsely represented in a proper

basis. Then compressive sensing technique is then used for multi-cluster C-RAN system to improve the fronthaul transfer efficiency.

2.5.2 Compression in the Downlink

In the downlink, the BBU pool first pre-codes messages for each user, and then compresses the pre-coded signals separately or jointly before transmitting to the RRHs. Then the RRHs decompress the received signals and forward them to the targeted users. The state of the art of literatures in terms of downlink fronthaul compression is summarised in Table 2.2.

Table 2.2: Summary of literature related to downlink fronthaul compression

Compression approach	Literature	System model	Main contribution
Joint compression (multivariate compression)	[22]	One cluster consists of multiple multi-antenna RRHs and multi-antenna users with limited per RRH fronthaul capacity	Alleviate the effect of additive quantisation noise at users by using joint compression
	[23]	Multiple interfering clusters each consist of multiple multi-antenna RRHs and multi-antenna user with limited per RRH fronthaul capacity	Propose a joint design of precoding and fronthaul compression for a multi-cluster scenario
	[24]	Multiple multi-antenna RRHs and multi-antenna user model with limited per RRH fronthaul capacity	Propose a method combining precoding, multivariate compression and RRH-user selection to maximise the energy efficiency
Point-to-point compression	[25]	Multiple multi-antenna RRHs and multi-antenna user model with limited per RRH fronthaul capacity	Propose a joint design of precoding and compression for downlink C-RAN which considers block-ergodic fading.

As the counterpart of the distributed source coding scheme for uplink C-RAN, a novel joint compression method, which is referred to as multivariate compression, is proposed for downlink C-RAN in [22]. Furthermore, a joint precoding and multivariate compression scheme is designed to solve the optimisation problem of maximising the weighted sum rate under power and fronthaul constraints. Besides, an iterative algorithm is developed to find a stationary point of the above problem. Numerical results demonstrate that the proposed joint design outperforms the traditional strategy which accomplishes precoding and compression separately. Since [22] focusses only on optimisation of precoding and compression within one cluster of RRHs without considering the influence

of inter-cluster interference, the proposed approach becomes less optimal when multiple mutually interfering clusters are involved. Based on this observation, an improved joint design of precoding and multivariate compression for inter-cluster C-RAN system is proposed in [23]. An iterative algorithm is developed to deal with the problem of maximising weighted sum rate. It is observed from the numerical results that the improved approach outperforms the existing methods based on intra-cluster optimisation. In [24], a joint design of precoding, multivariate compression and RRH-user selection is proposed for maximising the energy efficiency of downlink C-RAN system. With some novel relaxation and transformation techniques, the optimisation problem is converted into a continuous program. Then the authors propose an iterative method based on difference of convex algorithms to solve the relaxed problem. The numerical results shows that the proposed method can effectively improve the energy efficiency.

In addition to the aforementioned joint compression based methods, the authors in [25] investigate the joint precoding and fronthaul compression scheme for a practical downlink C-RAN scenario which incorporates block-ergodic fading with either instantaneous or stochastic CSI. Moreover, two types of BBU-RRH functional splitting at PHY, including compression-after-precoding (CAP) and compression-before-precoding (CBP), are introduced. For the former, BBU pool accomplishes all the baseband processing, while for the latter, BBU pool only calculates the precoding matrices and transmits the data and precoding matrices to the RRHs, then RRHs will complete the channel encoding and precoding. The optimisation problem of ergodic capacity maximisation for both CAP and CBP is solved by an efficient algorithm based on stochastic successive upper bound minimisation technology. The authors observe from the simulation results that the optimal BBU-RRH functional split for C-RAN system is determined by the interaction between the enhanced interference management abilities of CAP and the lower fronthaul requirements of CBP.

In contrast with the above fronthaul compression strategies that target at either the uplink or the downlink of C-RAN systems, joint design of fronthaul compression and precoding is proposed for a fronthaul-constrained full-duplex (FD) C-RAN system in [26]. Taking both the uplink-to-downlink and downlink-to-uplink interference into account, each RRH performs FD transmission with analog self-interference cancellation (SIC) and the BBU pool implements the cooperative interference management. Moreover, an additional digital SIC is executed at each RRH to further mitigate the residual self-interference

caused by FD transmission, which consequently lessens the heavy burden of the uplink fronthaul capacity.

Chapter 3

Theory of Compressive Sensing

Contents

3.1 Introduction **22**

3.2 Basic Concepts **24**

3.3 Robustness of Compressive Sensing **26**

3.4 Recovery Algorithms **29**

In this chapter, a promising technology called compressive sensing will be described in detail. First, an overview of compressive sensing along with several notable applications are presented. Then we will introduce the standard compressive sensing model and important notions, including sparsity and incoherence. A more practical compressive sensing model that considers data corruption will be further discussed. Following this, the key notion of the restricted isometry property (RIP) which is proved to be essential for studying the robustness of compressive sensing is introduced and different types of efficient measurement matrices are presented. Finally, a variety of popular recovery algorithms as well as the comparison among these algorithms is illustrated.

3.1 Introduction

Conventional methods of signal sampling follow the principle of the Nyquist-Shannon sampling theory which states that the sampling rate must be at least twice the maximum signal frequency (the so-called Nyquist rate) so that the original signal can be reconstructed without distortion. In fact, however, perfect reconstruction can still be possible even if the sampling criterion is not satisfied. Compressive sensing, also known as compressive sampling or compressed sensing, is a novel paradigm making this possible, as it is observed that many real-world signals are sparse or compressible, which enables one to reconstruct the signals from far fewer samples (or measurements) than conventional methods use.

The basic concept of compressive sensing was first proposed by Emmanuel Candes, Justin Romberg, Terence Tao and David Donoho in around 2004. However their papers actually appeared in print in 2006 [27–29] and inspired a mass of research activities in compressive sensing that mostly lies at the intersection of applied mathematics, electronic engineering, computer science and geophysics. An intrinsic difference between traditional Nyquist-Shannon sampling theory and compressive sensing is that compressive sensing performs data acquisition and compression simultaneously, rather than separately (thus the name). This relies on two key principles: sparsity, which refers to the characteristics of signals, and incoherence, which refers to the sensing modality [30]. Further details of these two notions will be given in Section 3.2.2.

With the knowledge that a signal is sparse or compressible, and that the coherence between representation basis and sensing basis is small, compressive sensing requires fewer measurements to perfectly recover the original signal than traditional approaches. It means that substantial reduction of data storage and computational resources can be achieved, which has certainly produced a large number of potential applications.

One of the notable instances of compressive sensing implementation is the single-pixel camera developed by Baraniuk and coworkers [31]. In contrast to conventional cameras which use charged-coupled device (CCD) or complementary metal-oxide-semiconductor (CMOS) technology, the single-pixel camera computes the incoherent measurements by using a digital micromirror device, where each measurement is a random sum of pixel

values taken across the whole image, and recovers images via sparse recovery algorithms. This new design accomplishes a simpler, smaller, and less-expensive camera that has overwhelming capability to flexibly and efficiently address high-dimensional data sets from video, hyperspectral imaging or other applications. As another implementation, magnetic resonance imaging (MRI) can also adopt compressive sensing to significantly reduce scan time and improve the visual quality [32]. Because it is evident, though implicit, that magnetic resonance images like angiography are often sparse or highly compressible.

In addition to the application of the single-pixel camera and MRI that fall into the class of image processing, there are also many signal processing applications of compressive sensing, such as channel coding (error correction) [33], channel estimation [34], etc. In [33], the authors show that efficient error correction can be achieved via l_1 minimisation, which is actually related to the problem of compressive sensing, namely, recovering signals from highly incomplete measurements. In [34], a notion of multipath sparsity is introduced. Exploiting this sparsity, the authors propose a new method to perform channel estimation based on the theory of compressive sensing. Further practical developments extend to other fields, while the theoretical framework of compressive sensing is enriched.

As mentioned in Section 2.5, another feasible application is the compressive sensing based cloud radio access network (C-RAN) system. Since in compressive sensing, only a simple operation (i.e., linear projection) is needed for collecting measurements, and relatively high computational requirements for reconstructing the original signals, making it feasible for a fully-centralised C-RAN network that has distributed remote radio heads (RRHs) with relatively simple functionalities to complete the linear operations, along with the baseband unit (BBU) pool equipped with high-performance processors to handle the signal reconstruction.

3.2 Basic Concepts

3.2.1 Standard Compressive Sensing Model

In engineering one often encounters problems in recovering a signal from the measured information. Mathematically speaking, consider a matrix $\Phi \in \mathbb{C}^{m\times n}$ with $m < n$ which stands for the linear measurement processing, and then the measurements $\mathbf{y} \in \mathbb{C}^{m\times 1}$ and the signal $\mathbf{b} \in \mathbb{C}^{n\times 1}$ of interest are connected through

$$\mathbf{y} = \Phi\mathbf{b}. \tag{3.1}$$

It has been proved that this underdetermined linear system of equations has infinitely many solutions. However, under certain assumptions it will surprisingly be feasible to find the solution [35]. Compressive sensing is the approach that makes this possible.

In general, finding the solution to (3.1) via compressive sensing is accomplished in three steps, including sparse representation, incoherent sampling (sensing), and sparse reconstruction, which will be introduced in the following sections respectively.

3.2.2 Fundamental Principles

This section introduces the two important principles that make search for solutions to (3.1) feasible: sparsity and incoherence.

Sparsity

Many natural signals have the feature that the signal itself is sparse or it may have sparsity in some transform domain. Consider a concrete example: a vector $\mathbf{b} \in \mathbb{C}^{n\times 1}$ is the object of interest, which can be expanded in an orthonormal basis $[\psi_1\psi_2\cdots\psi_n]$ as follows:

$$\mathbf{b} = \sum_{i=1}^{n} \psi_i x_i, \tag{3.2}$$

where $\{x_i \colon i = 1, 2, \cdots, n\}$ is the coefficient sequence of $\mathbf{b}$, and the orthonormal basis can be, for example, a Fourier basis, a wavelet basis, etc. Furthermore, (3.2) can be equivalently written as

$$\mathbf{b} = \Psi \mathbf{x}, \tag{3.3}$$

where Ψ is the $n \times n$ matrix with columns $\psi_1, \psi_2, \cdots, \psi_n$, and $\mathbf{x}$ is the coefficient vector. The object $\mathbf{b}$ is called s-sparse if $\mathbf{x}$ has at most s non-zero elements where s is small compared to n. This representation has clearly shown the underlying significance of sparsity: one can discard a large fraction of the coefficients without losing much information when the signal is sparse in some transform domain, which leads to more efficient data acquisition.

Incoherence

The sensing procedure and recovery process in compressive sensing pertain to a pair of bases (Φ, Ψ). The first basis Φ with l_2-normalised columns is so-called measurement basis which is used for sensing $\mathbf{b}$ in (3.1), and the second basis Ψ with orthonormal columns is so-called sparsity basis which is used to represent data $\mathbf{b}$.

The mutual coherence between the measurement basis and sparse basis is defined as follows [36]:

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \leq k, j \leq n} |\langle \phi_k, \psi_j \rangle|. \tag{3.4}$$

By definition, the coherence measures the maximum correlation between any two entries of Φ and Ψ, and the value is bounded in $[1, \sqrt{n}]$ according to linear algebra [30]. The incoherence between measurement basis Φ and sparsity basis Ψ is the key premise that guarantees the success of recovery algorithms. In short, given a fixed number of measurements, the smaller the coherence, the fewer the measurements that are needed and the larger the volume of effective information extracted from the original signal, and the better the recovery performance.

Various pairs of bases have been verified to be highly incoherent. For example, one can choose Φ as the spike basis and Ψ as the Fourier basis, then a signal which is most sparsely represented as a sum of sinusoids is measured using spikes. In other words, this sampling

process corresponds to the classical sampling in time or frequency domain, and therefore, the maximal incoherence (i.e., $\mu(\Phi, \Psi) = 1$) will be achieved. The pair of the identity and the Hadamard can also reach the lower bound [30, 36]. Moreover, it has been proved that random matrices (chosen as a measurement basis Φ) are highly incoherent with any fixed sparsity basis Ψ [30], especially the subgaussian random matrices whose entries are drawn independently from a subgaussian distribution. The definition of subgaussian is given as following:

Formally, a random variable X is called subgaussian if there exist positive constants C, k such that for all $t > 0$,

$$P(|X| \geq t) \leq Ce^{-kt^2}. \tag{3.5}$$

In this thesis, a subgaussian matrix will be adopted as a part of measurement matrix, which will be described in detail in the following chapter.

3.3 Robustness of Compressive Sensing

In practice, data corruption usually occurs in the process of data acquisition. Taking errors into account, a practical compressive sensing model can be expressed as

$$\begin{aligned} \mathbf{y} &= \Phi\mathbf{b} + \mathbf{e} \\ &= \Phi\Psi\mathbf{x} + \mathbf{e} \\ &= \mathbf{A}\mathbf{x} + \mathbf{e}, \end{aligned} \tag{3.6}$$

where $\mathbf{e}$ denotes the measurement error, and $\mathbf{A} = \Phi\Psi$ is called sensing matrix. In other words, $\mathbf{y}$ can be regarded as the measurements of $\mathbf{x}$ with regard to the "measurement matrix" $\mathbf{A}$. The following discussion will mainly concerned on this model bearing in mind that $\mathbf{x}$ may be a sparse representation of the signal of interest.

3.3.1 Restricted Isometry Property

In addition to the incoherence, the restricted isometry property (RIP) is another powerful tool to assess the quality of a measurement matrix, and it is also the key to study the robustness of compressive sensing. This notion was firstly introduced by Emmanuel Candes and Terence Tao in [37], named uniform uncertainty principle (UUP), and later defined as RIP in [33]. To start with, the restricted isometry constant is introduced.

Definition 3.1 [33]: For each integer $s = 1, 2, \cdots$, define the restricted isometry constant δ_s of a matrix $\mathbf{A} \in \mathbb{C}^{m \times n} (m < n)$ to be the smallest quantity such that

$$(1 - \delta_s)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_s)\|\mathbf{x}\|_2^2 \tag{3.7}$$

holds for all s-sparse vectors $\mathbf{x}$. Then $\mathbf{A}$ is said to obey the s-th order RIP with a constant δ_s. Note that this definition is only informative for $0 \leq \delta_s < 1$. The core idea of the above definition is that every subset T with s columns chosen from $\mathbf{A}$ is in fact near-orthogonal.

To see the relevance of RIP to the compressive sensing problem, consider the following case. Suppose that $\delta_{2s} < 1$, and that there are two distinct sparse solutions to the s-sparse signal $\mathbf{x}$, i.e., $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_2$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are supported on two distinct subsets T_1 and T_2 respectively ($\|T_1\|_2, \|T_2\|_2 \leq s$). Then

$$\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2) = 0. \tag{3.8}$$

Note that $(\mathbf{x}_1 - \mathbf{x}_2)$ is supported on $T_1 \cup T_2 (\|T_1 \cup T_2\|_2 \leq 2s)$.Then the observation that $\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 = 0$ can be easily derived by taking the l_2-norm of both sides in (3.8) and applying (3.7) with the assumption $\delta_{2s} < 1$, which obviously contradicts the hypothesis that the two solutions were distinct.

In other words, the assumption $\delta_{2s} < 1$ implies that distinct s-sparse vectors must have distinct measurement vectors. Consequently, this ensures the existence of efficient algorithms for seeking the unique solutions to the s-sparse signals based on their compressive measurements.

3.3.2 Potential Measurement Matrices

In fact, a variety of matrices are shown to satisfy the RIP condition and fulfil signal recovery with overwhelming probability [27, 28, 35, 38, 39]. Several notable measurement matrices are listed as follows:

- Random matrices

  This variety of matrices can be classified into unstructured random matrices and structured random matrices. The former mainly includes Gaussian random matrices, symmetric Bernoulli random matrices, and other subgaussian random matrices (Gaussian and Bernoulli random matrices are actually subgaussian [35]). The independent and identically distributed (i.i.d.) entries of these matrices are drawn from coresponding distributions respectively. It is proved that when the number of measurements $m \geq C \cdot s \ln(n/s)$ (where C is a constant depending on each instance), these random matrices will satisfy the RIP condition with high probability [28]. The subgaussian random matrix is popular because it is shown to be largely incoherent with most of sparsity bases, which benefits the later signal recovery. However, it has an obvious disadvantage that it will cost large computational resources to form this unstructured matrix which will also occupy large storage resources, making it difficult to implement in hardware. On the other hand, the structured random matrix, such as the partial Fourier matrix which is formed by taking random samples from Fourier transform [27], has the advantage of fast generation and easy storage of matrix. But such a matrix has the drawback that the entries are not independent and may be unable to guarantee low incoherence with sparsity bases, which may lead to a limitation of applications.

- Deterministic matrices

  The entries of deterministic matrices are drawn from certain distributions or sequences, such as the Toeplitz matrices [38], polynomial matrices [39] and so on. These matrices have the advantages of low construction complexity and easy hardware implementation, but the reconstruction accuracy will be lower compared to the random matrices.

In addition to above matrices, researchers have made an effort to design effective ma-

trices for better supporting the signal acquisition and recovery procedure. However how to design an optimal measurement matrix is still an open issue.

3.4 Recovery Algorithms

The popular recovery algorithms in compressive sensing can be divided into three categories: convex optimisation methods, greedy algorithms and thresholding algorithms. In this thesis, we mainly focus on the convex optimisation methods and greedy algorithms, which will be introduced in detail below.

3.4.1 Basis Pursuit Denoising

The problem of recovering the data $\mathbf{x}$ in (3.6) from corrupted measurements $\mathbf{y}$ can be translated into a convex optimisation problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimise}} \quad & \|\mathbf{x}\|_1 \\ \text{subject to} \quad & \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon, \end{aligned} \tag{3.9}$$

where ϵ is the noise level (i.e., $\|\mathbf{e}\|_2 \leq \epsilon$). This problem is known in many compressive sensing literature as basis pursuit denoising (BPDN) [40].

Clearly, this is a second-order cone program [41] which can be solved effectively by using various software packages, like CVX developed by Stephen Boyd and his students, l_1-magic developed by Candes and Romberg, etc. CVX is also the toolbox used for the signal recovery scheme proposed in this thesis. The running time of BPDN based on interior point methods is $\mathcal{O}(m^2 n^{3/2})$ [43], which is a considerable computational complexity compared with the greedy algorithms.

Robustness of signal recovery via BPDN

Theorem 3.1 [42]: Assume that $\delta_{2s} \leq \sqrt{2} - 1$, then the solution $\tilde{\mathbf{x}}$ to (3.9) obeys

$$\|\tilde{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 \cdot \epsilon + \frac{C_1 \cdot \|\mathbf{x} - \mathbf{x}_s\|_1}{\sqrt{s}} \tag{3.10}$$

for some constants C_0 and C_1, where $\mathbf{x}_s$ represents the vector with the s-largest values of $\mathbf{x}$ and set the rest $(n - s)$ values to zero.

As this theorem states, the recovery error is bounded by the sum of two terms. The first term is proportional to the noise level, and the second term is the error that occurs in the noiseless case. This theorem also suggests that with well-designed measurement matrices that obey the RIP condition, BPDN can guarantee the robust recovery of signals.

3.4.2 Greedy Algorithms

Three classical greedy algorithms, namely orthogonal matching pursuit (OMP) [44], compressive sampling matching pursuit (CoSaMP) [47] and sparsity adaptive matching pursuit (SAMP) [51] are introduced in this section. They are called greedy because at the selection process, one or several elements will be chosen according to certain criteria to update the support set at each iteration.

Orthogonal Matching Pursuit (OMP)

A formal description of OMP is given as follows [44].

Task: Find a sparse approximation of the solution to (3.9).

Input:

- The $m \times n$ measurement matrix $\mathbf{A}$
- The $m \times 1$ data vector (measurements) $\mathbf{y}$

- The threshold ϵ
- The sparsity level s of the signal of interest

Initialisation:

- The initial residual $\mathbf{r}_0 = \mathbf{y}$
- The initial signal estimate $\tilde{\mathbf{x}} = \mathbf{0}$
- The initial index set $S_0 = \emptyset$
- The initial matrix with chosen columns $\mathbf{A}_0 = \emptyset$
- The iteration counter $t = 1$

Iteration:

1) Find the index s_t by solving the problem $s_t = \arg\max_{j=1,2,\cdots,n} |\langle \mathbf{r}_{t-1}, \mathbf{a}_j \rangle|$.
2) Update the index set $S_t = S_{t-1} \cup \{s_t\}$ and the matrix of chosen columns $\mathbf{A}_t = \mathbf{A}_{t-1} \cup \mathbf{a}_{s_t}$.
3) Update the signal estimate by solving a least squares problem $\tilde{\mathbf{x}}_t = \arg\min_{\tilde{\mathbf{x}}_t} \|\mathbf{y} - \mathbf{A}_t\tilde{\mathbf{x}}_t\|_2 = (\mathbf{A}_t^T\mathbf{A}_t)^{-1}\mathbf{A}_t^T\mathbf{y}$.
4) Refine the residual $\mathbf{r}_t = \mathbf{y} - \mathbf{A}_t\tilde{\mathbf{x}}_t$.
5) If $\|\mathbf{r}_t\|_2 \leq \epsilon$, stop. Otherwise, increment t and return to Step 1 until $t > s$.

Output:

- The estimated signal $\tilde{\mathbf{x}}_t$
- The index set S_t

Note that in the stage of updating the signal estimate, the solution is actually obtained by the following equation

$$\mathbf{A}_t^T(\mathbf{A}_t\tilde{\mathbf{x}}_t - \mathbf{y}) = -\mathbf{A}_t^T\mathbf{r}_t = 0, \tag{3.11}$$

where $\mathbf{r}_t = \mathbf{y} - \mathbf{A}_t\tilde{\mathbf{x}}_t$ is the formula of the residual in Step 4. Above equation (3.11) indicates that the residual $\mathbf{r}_t$ is always orthogonal to the columns in $\mathbf{A}_t$, which in turn implies that the algorithm will always select a new column for the index set in the next iteration (and afterwards). The orthogonalisation between the residual $\mathbf{r}_t$ and the columns of matrix $\mathbf{A}_t$ is exactly the reason for the name of this algorithm – orthogonal matching pursuit.

The computational complexity of the OMP algorithm is $\mathcal{O}(mns)$, which is much lower than BPDN and make it easier to implement. Besides, OMP algorithm is shown to perform better when the measurement matrix is structured [45, 46]. But when it comes to unstructured matrices, like the Gaussian or Bernoulli measurement matrices, with high probability OMP fails to recover all the sparse signals [44].

Compressive Sampling Matching Pursuit (CoSaMP)

CoSaMP is recognised as an improved algorithm based on OMP. Unlike OMP, CoSaMP selects multiple indices at each iteration, which allows faster running time. The procedure of CoSaMP is described as follows [47].

Input:

- The $m \times n$ measurement matrix $\mathbf{A}$
- The $m \times 1$ data vector (measurements) $\mathbf{y}$
- The threshold ϵ
- The sparsity level s of the signal of interest

Initialisation:

- The initial residual $\mathbf{r}_0 = \mathbf{y}$
- The initial signal estimate $\tilde{\mathbf{x}} = \mathbf{0}$
- The initial index set $S_0 = \emptyset$
- The initial matrix with chosen columns $\mathbf{A}_0 = \emptyset$
- The iteration counter $t = 1$

Iteration:

1) Compute $\mathbf{p} = |\langle \mathbf{r}_{t-1}, \mathbf{a}_j \rangle| (j = 1, 2, \cdots, n)$ and select the largest $2s$ components of $\mathbf{p}$, then gather the $2s$ indices corresponding to the $2s$ largest components into a set J_0.
2) Update the index set $S_t = S_{t-1} \cup J_0$ and the matrix of chosen columns $\mathbf{A}_t = \mathbf{A}_{t-1} \cup \mathbf{a}_j (\text{for all} j \in J_0)$.
3) Find the solution $\tilde{\mathbf{x}}_t$ to the least squares problem $\tilde{\mathbf{x}}_t = \arg\min_{\tilde{\mathbf{x}}_t} \|\mathbf{y} - \mathbf{A}_t \tilde{\mathbf{x}}_t\|_2 = (\mathbf{A}_t^T \mathbf{A}_t)^{-1} \mathbf{A}_t^T \mathbf{y}$.
4) Select the s largest elements in $|\tilde{\mathbf{x}}_t|$ as $\tilde{\mathbf{x}}_{ts}$, and then update the index set $S_t = S_{ts}$ and the matrix $\mathbf{A}_t = \mathbf{A}_{ts}$.
5) Update the residual $\mathbf{r}_t = \mathbf{y} - \mathbf{A}_{ts} \tilde{\mathbf{x}}_{ts}$.
5) If $\|\mathbf{r}_t\|_2 \leq \epsilon$, stop. Otherwise, increment t and return to Step 1 until $t > s$.

Output:

- The estimated signal $\tilde{\mathbf{x}}_{ts}$
- The index set S_{ts}

As seen from above procedure, CoSaMP incorporates the idea of backtracking (Step 4), which leads to more stable signal recovery. The running time of CoSaMP is $\mathcal{O}(mn)$, which is significantly lower than OMP and BPDN.

In addition to OMP and CoSaMP, other popular greedy algorithms, including stage-wise orthogonal matching pursuit (StOMP) [48], regularised orthogonal matching pursuit (ROMP) [49], subspace pursuit (SP) [50] and other improved algorithms based on existing ones have been well-developed. The complexity of the StOMP and the ROMP is around $\mathcal{O}(mns)$, which is comparable with that of the OMP algorithm, while the SP only needs a running time of $\mathcal{O}(mn \ln s)$ [50]. It is worth pointing out that the SP algorithm is quite similar to the CoSaMP algorithm (though these works are developed by different teams independently at almost the same time). The main difference between these two algorithms lies in in the way of adding new atoms to the index set. More precisely, in each iteration, CoSaMP adds $2s$ candidates to the index set, while in the SP algorithm, only s components are selected. This certainly allows more efficient computation in the SP algorithm, but at the cost of higher complexity in terms of the underlying analysis.

Although the above algorithms are verified to have the capability of successfully recovering certain signals, all these algorithms have the obvious shortcoming that they assume that the sparsity of the signal is already known, whereas this information may not be available in many practical scenarios. To make up for this shortcoming, a more practical greedy algorithm is proposed, that is, sparsity adaptive matching pursuit (SAMP) [51].

Sparsity Adaptive Matching Pursuit (SAMP)

The details of the SAMP algorithm are presented as follows.

Input:

- The $m \times n$ measurement matrix $\mathbf{A}$
- The $m \times 1$ data vector (measurements) $\mathbf{y}$
- The threshold ϵ
- Step size t

Initialisation:

- The initial residual $\mathbf{r}_0 = \mathbf{y}$
- The initial signal estimate $\tilde{\mathbf{x}} = \mathbf{0}$
- The initial index set $\Lambda_0 = \emptyset$
- The iteration counter $k = 1$
- Size of index set $L = t$

Iteration:

1) Compute $\mathbf{p} = |\langle \mathbf{r}_{k-1}, \mathbf{a}_j \rangle| (j = 1, 2, \cdots, n)$, then gather the L indices corresponding to the L largest components of $\mathbf{p}$ into a set S_k.
2) Augment the candidate index set $C_k = \Lambda_{k-1} \cup S_k$;
3) Compute $\tilde{\mathbf{x}} = |\mathbf{A}_{C_k}^{\dagger}\mathbf{y}|$, and then select L indices corresponding to the L largest components of $\tilde{\mathbf{x}}$ to form the index set Λ.
4) Update the residual $\mathbf{r} = \mathbf{y} - \mathbf{A}_{\Lambda}\mathbf{A}_{\Lambda}^{\dagger}\mathbf{y}$.
5) If $\|\mathbf{r}\|_2 \leq \epsilon$, stop and output $\tilde{\mathbf{x}}_{\Lambda}$ and Λ; Otherwise, if $\|\mathbf{r}\|_2 \geq \|\mathbf{r}_{k-1}\|_2$, then update the size of index set $L = L + t$ and go back to Step 1; If the above two halting conditions are not true, increment k and refine $\Lambda_k = \Lambda$, $\mathbf{r}_k = \mathbf{r}$ and return to Step 1.

Output:

- The estimated signal $\tilde{\mathbf{x}}_{\Lambda} = \mathbf{A}_{\Lambda}^{\dagger}\mathbf{y}$
- The index set Λ

Note that the step size in the SAMP algorithm is required to be equal or less than the sparsity of the signal (i.e., $t \leq s$). This means that there is a trade-off between the step size and the recovery speed (as the larger step size allows fewer iterations). Clearly, setting the step size to one is the safest choice to avoid overestimation. It is also shown from the

empirical results that for signals with exponentially decayed magnitude, a small step size is preferable, while for binary sparse signals, a large step size is more appropriate [51].

An important feature of the SAMP is that it does not require the sparsity of the signal as a *prior* information, which makes it more appropriate for many practical scenarios where the sparsity is not available. Meanwhile, it can still be able to guarantee the stability of signal recovery.

In fact, SAMP can be regarded as a generalised version of the OMP and the SP. On the one hand, when the step size $t = 1$, SAMP can be roughly regarded as a generalised OMP algorithm associated with the feature that it can refine the index set by removing the bad candidates in each iteration. This offers a more accurate reconstruction than OMP despite using a few more iterations to achieve that accuracy. On the other hand, when the step size $t = s$, SAMP turns into the SP, in which case it only requires one iteration to find the sparse solution to the original signal. Consequently, the computational complexity of SAMP is bounded by $\mathcal{O}(mns)$ (the complexity of OMP), and may be further reduced to $\mathcal{O}(mn \ln s)$ (the complexity of SP). It is clear that the actual running time of SAMP depends on the number of iterations, which is further determined by the chosen step size. But how to design the optimal value of step size is still an open issue.

3.4.3 Comparison Between BPDN and Greedy Algorithms

In principle, all these algorithms work reasonably well in practice. Determining which one should be favoured is the matter of concrete situation, i.e., specific requirements such as the speed of the algorithm, the computational complexity, the recovery accuracy and so on.

As a first criterion, the minimal number of measurements required for reconstructing a sparse signal may vary with each algorithm. For example, the numerical results in [51] show that for Gaussian sparse signals, BPDN requires more measurements than one of the other greedy algorithms to achieve the same recovery accuracy, while for binary sparse signals, the opposite result is presented. The complexity of the algorithms may be a second criterion, and it has been shown in this section that on the whole greedy algorithms

run with lower computational complexity than BPDN. Specifically, among these greedy algorithms the CoSaMP algorithm offers the lowest complexity. Table 3.1 shows the comparison of these algorithms in terms of the order of the complextity [43, 47, 51].

Table 3.1: Comparison of complexity of different recovery algorithms

	BPDN	**OMP**	**CoSaMP**	**SP**	**SAMP** *
Complexity	$\mathcal{O}(m^2n^{3/2})$	$\mathcal{O}(mns)$	$\mathcal{O}(mn)$	$\mathcal{O}(mn\ln s)$	$[\mathcal{O}(mn\ln s), \mathcal{O}(mns)]$

* the specific computational complexity of SAMP is related to the step size t [51].

However, it is notable that only the BPDN and the SAMP algorithm do not require the sparsity of signal as an input parameter. It allows a more practical implementation in many applications, and this is also one of the reasons why BPDN and SAMP are adopted in the C-RAN system proposed in this thesis.

Chapter 4

Cell-free C-RAN System with Compressive Sensing

Contents

4.1 Introduction 38

4.2 Uplink C-RAN System Model 39

4.3 Proposed Compression and Recovery Scheme 41

4.4 System Analysis 48

4.5 Numerical Results and Discussion 49

4.6 Summary 55

4.1 Introduction

Many advanced techniques have been developed to address the challenge of the high fronthaul load in an uplink cloud radio access network (C-RAN) system. Compressive sensing, as mentioned in Chapter 2, is one promising technique to achieve efficient fronthaul compression and thus significantly reduce the fronthaul load. In this chapter, we are interested in applying this novel technology to an uplink C-RAN system with the aim of improving fronthaul performance. Specifically, a very large number of multi-antenna remote radio heads (RRHs) simultaneously serve a smaller number of single-antenna users

over a service area in which there are no cells or cell boundaries, and hence is decreibed as “cell-free” [52].

In particular, this work is envisaged to be applied to the internet of things (IoT) [53], or to massive machine-type communication (mMTC) [54] in 5G, in which the systems might not be fully loaded and the user data tends to be sparse. Moreover, the channels among the low-mobility devices can be assumed to be stable and known at each terminal [54]. With the assumption that the uplink signals are sparse and the channel state information is already known, compressive sensing can be easily adopted to process signals at each RRH and reconstruct user data at the BBU pool.

The rest of this chapter is organised as follows. Section 4.2 describes the uplink C-RAN system model. Section 4.3 develops the compression and recovery scheme for this system. In particular, two different approaches, i.e., the BPDN and the SAMP algorithms, are used to reconstruct the sparse signals at the BBU pool. The system analysis is given in Section 4.4, prior to the numerical results presented in Section 4.5. Finally, we summarise this chapter in Section 4.6.

4.2 Uplink C-RAN System Model

We consider an uplink C-RAN system with M RRHs where each RRH is equipped with N antennas and connected to the BBU pool via a fronthaul network. There are a total of K single-antenna users being served by all the RRHs. All the RRHs and users are randomly deployed in a large area, as depicted in Fig. 4.1. Denote the signal transmitted by the k-th user as x_k, $k = 1, 2, \ldots, K$. Then the received signal at the m-th RRH can be expressed as

$$\mathbf{y}_m = \sum_{k=1}^{K} \mathbf{g}_{mk} x_k + \mathbf{n}_m, \tag{4.1}$$

where $m \in \{1, \ldots, M\}$, $\mathbf{n}_m = [n_1\ n_2\ \cdots\ n_N]^T$ is the $N \times 1$ vector which represents the standard complex Gaussian noise at the m-th RRH, and $\mathbf{g}_{mk} \in \mathbb{C}^{N\times 1}$ is the channel vector between the k-th user and the m-th RRH. The following assumptions are made regarding the channel model:

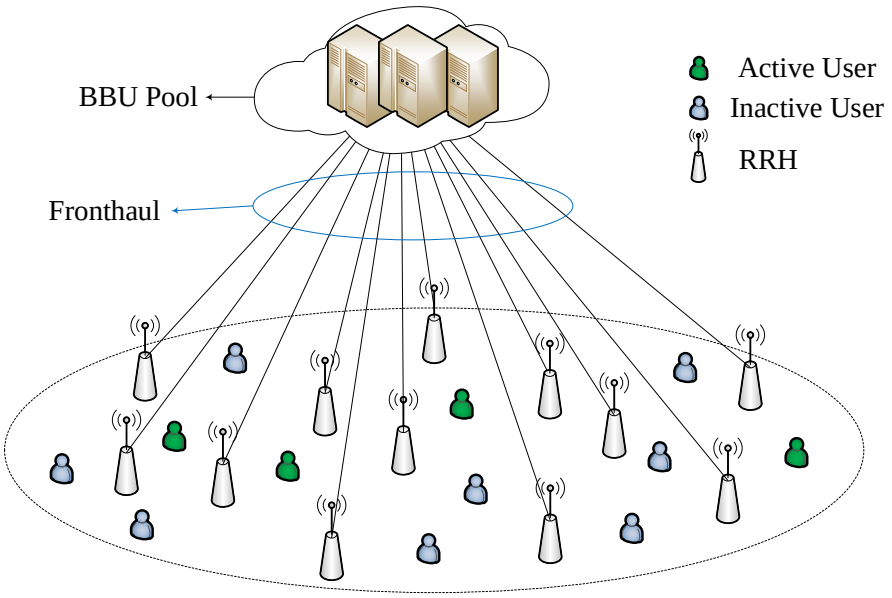


Figure 4.1: Uplink cell-free C-RAN system

- The channel model incorporates both the large scale fading and small scale fading, and the latter takes both path loss and shadowing into account.

- The channel $\mathbf{g}_{mk}$ can be modelled as

$$\mathbf{g}_{mk} = \beta_{mk}^{1/2}\mathbf{h}_{mk}, \tag{4.2}$$

  where β_{mk} denotes the large scale fading coefficient which is uncorrelated for different RRH-user pairs, and $\mathbf{h}_{mk} \in \mathbb{C}^{N\times 1}$ represents the small scale fading coefficients. β_{mk} is supposed to be normalised so that $\sum_{i=1}^{M}(\beta_{ik})^2 = M$. We assume that small scale fading coefficients are independent and identically distributed (i.i.d.) standard complex Gaussian variables with zero mean and unit variance.

- We assume that the fronthaul connecting the RRHs and the BBU pool is noiseless and can provide infinite capacity. Moreover, the BBU pool is assumed to have perfect knowledge of channel state information $\{\mathbf{g}_{mk} : \forall m, k\}$.

Concatenate the transmitted signals from all users as $\mathbf{x} = [x_1\ x_2\ \ldots\ x_K]^T$, and then

the received signal (4.1) can be re-written as

$$\mathbf{y}_m = \mathbf{G}_m\mathbf{x} + \mathbf{n}_m,\ m \in \{1, \ldots, M\} \tag{4.3}$$

where $\mathbf{G}_m \in \mathbb{C}^{N\times K}$ denotes the combined channel matrix from all users to the m-th RRH and is given by

$$\mathbf{G}_m = [\mathbf{g}_{m1}\ \mathbf{g}_{m2}\ \ldots\ \mathbf{g}_{mK}]. \tag{4.4}$$

Here, we should note the assumption that the transmitted data $\mathbf{x} = [x_1\ x_2\ \ldots\ x_K]^T$ is sparse. In practice, $\mathbf{x}$ might be sparse because of massive bursty traffic in the random access mode [53, 54]. For example, in a mMTC system, a massive number of MTC devices usually generate infrequent and small amounts of data [54]. Besides, these devices will access to the network in the random access mode, and the time interval for each MTC device's communications with the MTC server is strictly controlled. Such features can reduce peaks in the data traffic because the access of MTC devices within the same access duration is distributed across this interval [54]. Hence the transmitted signals tend to be sparse.

So in this thesis, we consider the scenario where uplink signals are sparse in which only a small proportion of users transmit at any given time, and we call a user *active* if it is transmitting non-zero data. Denote the set of non-zero elements of $\mathbf{x}$ as S, and $s \triangleq \#|S|(s \ll K)$ as the sparsity level, i.e., the number of active users. We assume $s \ll MN$ (as this "cell-free" system is defined, a large number of RRHs serve a smaller number of *active* users).

4.3 Proposed Compression and Recovery Scheme

By exploiting uplink signal sparsity, a compressive sensing based scheme will be applied to reduce the fronthaul load in this cell-free C-RAN system. The details of the proposed scheme are presented as follows.

4.3.1 Fronthaul Compression and Forward

Before transmitting to the BBU pool, the received signals need to be compressed at each RRH. This is called fronthaul compression, which can be achieved via a simple linear operation, namely, multiplying the received signal $\mathbf{y}_m$ with a compression matrix. As discussed in Chapter 3, subgaussian random matrices are verified to satisfy the RIP with high probability so that robust signal recovery can be realised. Therefore, we generate the compression matrix entries from subgaussian random variables as follows:

$$\mathbf{C}_m(r,n) = \sqrt{\frac{1}{MR}}\exp(j\alpha_{rn}), \tag{4.5}$$

where α_{rn} is i.i.d. uniformly distributed variable over $[0, 2\pi)$, and R ($R \leq N$) denotes the number of measurements at each RRH. The factor $\sqrt{\frac{1}{MR}}$ is for normalisation.

Adopting this matrix $\mathbf{C}_m \in \mathbb{C}^{R\times N}$ to compress the $(N \times 1)$ vector $\mathbf{y}_m$, and the $(R \times 1)$ compressed vector can be expressed as

$$\mathbf{f}_m = \mathbf{C}_m\mathbf{y}_m, \tag{4.6}$$

which is then transmitted to the BBU pool through a noiseless fronthaul link.

Note that the measurement matrices $\mathbf{C}_m$ are generated locally and hence are available at the BBU pool.

4.3.2 Signal Recovery at BBU Pool

The signal recovery procedure can be divided into three steps: compressive sensing reconstruction, active user detection and original signal recovery. Considering that the number of active users (the sparsity) is unknown to the BBU pool, the BPDN and SAMP algorithms are considered as two feasible approaches for signal recovery. Furthermore, the relatively high recovery accuracy of these two algorithms is another reason why they are adopted in this scenario.

Recovery with BPDN

The flow chart of recovery method using BPDN is shown in Fig. 4.2. The three steps shown in the chart are described in detail below.

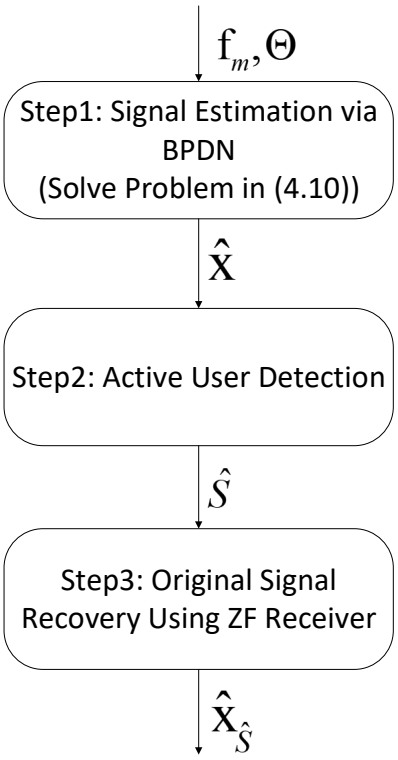


Figure 4.2: The flow chart of signal recovery using BPDN

1) Rough signal estimation using BPDN: Upon receiving the measurements (compressed signal) from each RRH, compressive sensing based recovery will be carried out to obtain a rough estimation of the user data. Firstly, concatenate the compressed vectors $\{\mathbf{f}_m : m = 1, 2, \ldots, M\}$ as $\mathbf{f} = [\mathbf{f}_1^T \ \mathbf{f}_2^T \ \ldots \ \mathbf{f}_M^T]^T \in \mathbb{C}^{MR\times 1}$. Substitute (4.3) into (4.6) and then (4.6) can be equivalently written as

$$\mathbf{f} = \Theta\mathbf{x} + \mathbf{n}, \tag{4.7}$$

where $\Theta \in \mathbb{C}^{MR\times K}$ is the aggregate measurement matrix which can be written as

$$\Theta = [(\mathbf{C}_1\mathbf{G}_1)^T \ (\mathbf{C}_2\mathbf{G}_2)^T \ \ldots \ (\mathbf{C}_M\mathbf{G}_M)^T], \tag{4.8}$$

and $\mathbf{n}$ is the aggregate noise

$$\mathbf{n} = [(\mathbf{C}_1\mathbf{n}_1)^T \ (\mathbf{C}_2\mathbf{n}_2)^T \ \dots \ (\mathbf{C}_M\mathbf{n}_M)^T] \in \mathbb{C}^{MR\times 1}. \tag{4.9}$$

Note that Θ is known at the BBU pool as a direct result of the fact that both $\{\mathbf{G}_m\}$ and $\{\mathbf{C}_m\}$ are available at the BBU pool.

Therefore, the problem of rough estimation of $\mathbf{x}$ can be formed as BPDN according to the theory of compressive sensing introduced in Chapter 3, as shown in (4.10), in which $\mathbf{f}$ is regarded as the measurements, Θ is the measurement matrix, $\mathbf{x}$ is the sparse signal need to be reconstructed and $\mathbf{n}$ is the measurement error.

$$\begin{aligned} \underset{\hat{\mathbf{x}}}{\text{minimise}} \quad & \|\hat{\mathbf{x}}\|_1 \\ \text{subject to} \quad & \|\Theta\hat{\mathbf{x}} - \mathbf{f}\|_2 \leq \lambda. \end{aligned} \tag{4.10}$$

where λ is the threshold parameter.

2) Active user detection: Our target is to find the set of active users $\hat{S}$ based on the estimated $\hat{\mathbf{x}}$. This can be achieved by the following strategy:

- Firstly, sort $\{|\hat{\mathbf{x}}(k)| : k = 1, 2, \dots, K\}$ in descending order and get the corresponding indices $\{k_1, k_2, \dots, k_K\}$;
- Secondly, initialise $i = 1$ and $\hat{S} = \{k_1\}$;
- Compute

$$u = \left\| \mathbf{f} - \Theta_{\hat{S}}\Theta_{\hat{S}}^{\dagger}\mathbf{f} \right\|_2, \tag{4.11}$$

  where $\Theta_{\hat{S}}^{\dagger}$ denotes the Moore-Penrose pseudoinverse of $\Theta_{\hat{S}}$ and is obtained by

$$\Theta_{\hat{S}}^{\dagger} = (\Theta_{\hat{S}}^H\Theta_{\hat{S}})^{-1}\Theta_{\hat{S}}^H. \tag{4.12}$$

  If $u > \lambda$, then update $i = i + 1$, $\hat{S} = \hat{S} \cup \{k_i\}$ and repeat until $u \leq \lambda$ or $i = MR$ (this is because we can reconstruct MR unknowns at most as we only have MR measurements in total at the BBU pool).

- Finally output $\hat{S}$.

3) Original signal recovery: The BBU pool adopts the zero-forcing (ZF) technique [55] to recover user data $\hat{\mathbf{x}}_{\hat{S}}$ based on the obtained $\hat{S}$, i.e., $\hat{\mathbf{x}}_{\hat{S}} = (\Theta_{\hat{S}}^{\dagger})\mathbf{f}$, $\hat{\mathbf{x}}_{\{1,\ldots,K\}\setminus\hat{S}} = \mathbf{0}$.

Clearly, solving the problem in (4.10) that estimates the sparse signal $\hat{\mathbf{x}}$ is a conventional type of convex relaxation approach [41]. It has the advantage of high recovery accuracy and stability, but at the cost of high computational complexity, around $\mathcal{O}((MR)^2 N^{3/2})$ (using interior point method).

Recovery with SAMP

Considering lower computational complexity, the SAMP algorithm can be an alternative method to achieve signal reconstruction. The description of the SAMP algorithm is given as follows.

Input:

- The $MR \times K$ measurement matrix Θ
- The $MR \times 1$ compressed vector (measurements) $\mathbf{f}$
- The threshold λ
- Step size t

Initialisation:

- The initial residual $\mathbf{r}_0 = \mathbf{f}$
- The initial signal estimate $\hat{\mathbf{x}} = \mathbf{0}$
- The initial index set $\hat{S}_0 = \emptyset$
- The iteration counter $l = 1$

- Size of index set $L = t$

Iteration:

1) Compute $\mathbf{p} = |\langle \mathbf{r}_{l-1}, \theta_j \rangle| (j = 1, 2, \cdots, K)$, then gather the L indices corresponding to the L largest components of $\mathbf{p}$ into a set J_l.

2) Augment the candidate index set $C_l = \hat{S}_{l-1} \cup J_l$;

3) Compute $\hat{\mathbf{x}} = |\Theta_{C_l}^{\dagger}\mathbf{f}|$, and then select L indices corresponding to the L largest components of $\hat{\mathbf{x}}$ to form the index set $\hat{S}$.

4) Update the residual $\mathbf{r} = \mathbf{f} - \Theta_{\hat{S}}\Theta_{\hat{S}}^{\dagger}\mathbf{f}$.

5) If $\|\mathbf{r}\|_2 \leq \lambda$, stop and output $\hat{\mathbf{x}} = \Theta_{\hat{S}}^{\dagger}\mathbf{f}$ and $\hat{S}$; Otherwise, if $\|\mathbf{r}\|_2 \geq \|\mathbf{r}_{l-1}\|_2$, then update the size of index set $L = L + t$ and go back to Step 1; If the above two halting conditions are not true, increment l and refine $\hat{S}_l = \hat{S}$, $\mathbf{r}_l = \mathbf{r}$ and return to Step 1.

Output:

- The estimated signal $\hat{\mathbf{x}}$
- The index set $\hat{S}$

After obtaining the estimated signal $\hat{\mathbf{x}}$ and the set of active users $\hat{S}$, the BBU pool uses the same ZF receiver to recover the original data from users $\hat{\mathbf{x}}_{\hat{S}} = (\Theta_{\hat{S}}^{\dagger})\mathbf{f}$, $\hat{\mathbf{x}}_{\{1,\ldots,K\}\setminus\hat{S}} = \mathbf{0}$.

It is evident that SAMP can estimate the sparse signal $\hat{\mathbf{x}}$ and detect the set of active users $\hat{S}$ simultaneously, which is relatively simpler in contrast with BPDN. Moreover, unlike other greedy algorithms like OMP, CoSaMP, SP, etc., SAMP does not require the sparsity of signal as an input parameter, which makes it more feasible for this scenario where the number of active users is unknown.

4.3.3 Overall Compression and Recovery Scheme

As illustrated in Fig. 4.3, the overall compression and recovery scheme for the uplink of cell-free C-RAN system are summarised in three steps:

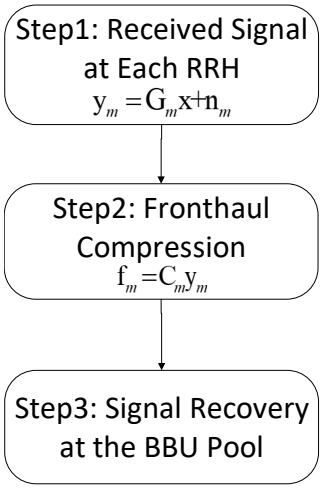


Figure 4.3: The overall compression and recovery scheme

- **Step 1** (Signals Received at RRHs): The signals from all users received at the m-th RRH $\mathbf{y}_m = \mathbf{G}_m\mathbf{x} + \mathbf{n}_m$, as in (4.3).

- **Step 2** (Fronthaul Compression and Forward): The m-th RRH uses a local compression matrix $\mathbf{C}_m \in \mathbb{C}^{R\times N}$ to compress the received signal $\mathbf{y}_m \in \mathbb{C}^{N\times 1}$ into a lower dimensional vector $\mathbf{f}_m \in \mathbb{C}^{R\times 1}$, as in (4.6). Then the m-th RRH transmits the compressed vector to the BBU pool via the fronthaul link.

- **Step 3** (Signal Recovery at the BBU Pool): The BBU pool gathers the compressed data from all RRHs as $\mathbf{f} = [\mathbf{f}_1^T\ \mathbf{f}_2^T\ \ldots\ \mathbf{f}_M^T]^T$ and then performs signal reconstruction using either BPDN or SAMP algorithm to estimate the transmitted data $\hat{\mathbf{x}}$ and the set of active users $\hat{S}$. After this, ZF receiver is applied to recover the transmitted signals $\hat{\mathbf{x}}_{\hat{S}}$.

4.4 System Analysis

RIP is a crucial feature that confirms whether it is feasible to achieve robust compressive sensing recovery using the measurement matrix. The proof of the RIP for a matrix such as occurs in this system, which is an aggregation of the channel matrix $\{\mathbf{G}_m\}$ and the compression matrix $\{\mathbf{C}_m\}$, is complicated and is beyond the scope of the present work. However, motivated by the proposed theorem and the techniques of proving RIP in [20], we argue that the RIP condition will also hold for our aggregated measurement matrix Θ. This argument is also supported by the later simulation results.

Theorem 4.1 (Stability of BPDN recovery under RIP [20]): Assume that the threshold parameter λ in (4.10) satisfies $\lambda \geq \|\mathbf{n}\|_2$: if the measurement matrix Θ obeys the $2s$-order RIP and the restricted isometry constant δ satisfies $\delta \leq \sqrt{2}-1$, then the solution $\hat{\mathbf{x}}$ to (4.10) satisfies

$$\|\mathbf{x}-\hat{\mathbf{x}}\|_2 \leq c_1\lambda, \tag{4.13}$$

where $c_1 = 4\frac{\sqrt{1+\delta}}{1-(1+\sqrt{2})\delta}$.

Based on Theorem 1, the conditions for accurate active user detection can be obtained as follows.

Theorem 4.2 (Conditions for Correct Active User Detection [20]): Assume that λ in (4.10) satisfies $\lambda \geq \|\mathbf{n}\|_2$ and the measurement matrix Θ obeys the $2s$-order RIP with the restricted isometry constant $\delta \leq \sqrt{2}-1$. If

$$|\mathbf{x}(i)| > \sqrt{2}c_1\lambda, \forall i \in S$$

is true, where c_1 is the same constant in Theorem 1, then

1) the estimated $\hat{S}$ is correct, i.e., $\hat{S} = S$.

2) the reconstructed signal $\hat{\mathbf{x}}$ from ZF receiver satisfies $\|\mathbf{x}-\hat{\mathbf{x}}\|_2 \leq \frac{1}{\sqrt{1-\delta}}\|\mathbf{n}\|_2$.

Theorem 4.1 and Theorem 4.2 provide the theoretical conditions for the correct active user detection and the performance result of signal recovery using BPDN. The following

theorem is given as the performance analysis of SAMP for sparse signal recovery in cell-free C-RAN system.

Theorem 4.3 (Stability of SAMP recovery under RIP [51]): Denote $s_t = t\lceil s/t \rceil$, where t is the step size defined in SAMP algorithm. Denote the energy (l_2 norm) of the noise as ϵ. If measurement matrix Θ satisfies the RIP with $\delta_{3s_t} < 0.06$, then the reconstructed signal $\hat{\mathbf{x}}$ of the SAMP algorithm is within the noise level

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq c_2\sqrt{\epsilon}, \tag{4.14}$$

where $c_2 = (1 + \delta_{3s_t})/(\delta_{3s_t}(1 - \delta_{3s_t}))$.

The proof of the above theorems can be found in [20] and [51] respectively, which further shows that these theorems can provide sufficient conditions of compressive sensing recovery algorithms for guarantee of exact reconstruction.

4.5 Numerical Results and Discussion

In this section, the performance of proposed compressive sensing based compression and signal reconstruction scheme in the cell-free C-RAN system will be illustrated.

4.5.1 Parameter Setup

We consider a C-RAN system with M RRHs where each RRH is equipped with $N = 8$ antennas, and a total of $K = 500$ single-antenna users. All users and RRHs are randomly and evenly distributed in a square area of 1×1 km. We use parameter settings resembling those in [52] to establish our channel model. Suppose the small scale fading $\mathbf{h}_{mk} \sim \mathcal{CN}(0, 1)$ and the large scale fading parameter is modelled as

$$\beta_{mk} = 10^{\frac{\mathrm{PL}_{mk}}{10}} \cdot \mathrm{Z}_{mk}, \tag{4.15}$$

where $\mathrm{Z}_{mk} = 10^{\frac{\sigma_{sh} z_{mk}}{10}}$ denotes the shadowing with standard deviation $\sigma_{sh} = 8$ dB and $z_{mk} \sim \mathcal{N}(0,1)$, and PL_{mk} represents the path loss coefficient (in dB) which is given by

$$\mathrm{PL}_{mk} = \begin{cases} -L - 35\log_{10}(d_{mk}), \text{ if } d_{mk} > d_1 \\ -L - 15\log_{10}(d_1) - 20\log_{10}(d_{mk}), \\ \qquad\qquad \text{if } d_0 < d_{mk} \leq d_1 \\ -L - 15\log_{10}(d_1) - 20\log_{10}(d_0), \text{ if } d_{mk} \leq d_0 \end{cases} \tag{4.16}$$

where $L = 140.7$ dB, $d_0 = 10$ m, and $d_1 = 50$ m.

Denote the number of active users (sparsity) as s and the number of measurements at each RRH as R. A binary sparse signal with equal transmit SNR is assumed in all examples. The threshold parameter λ in (4.10) is set to $\sqrt{2N}$, which will satisfy the condition required in Theorem 1 and hence achieving robust BPDN recovery. Moreover, we adopt the step size $t = 1, 5, 10$ in the SAMP algorithm.

Denote the number of accurately detected active users as q, then the percentage of recovered signals is defined to be the expected value of q/s, which is given by

$$P_{\text{recovered}} \triangleq E\left(\frac{q}{s}\right) \times 100\%. \tag{4.17}$$

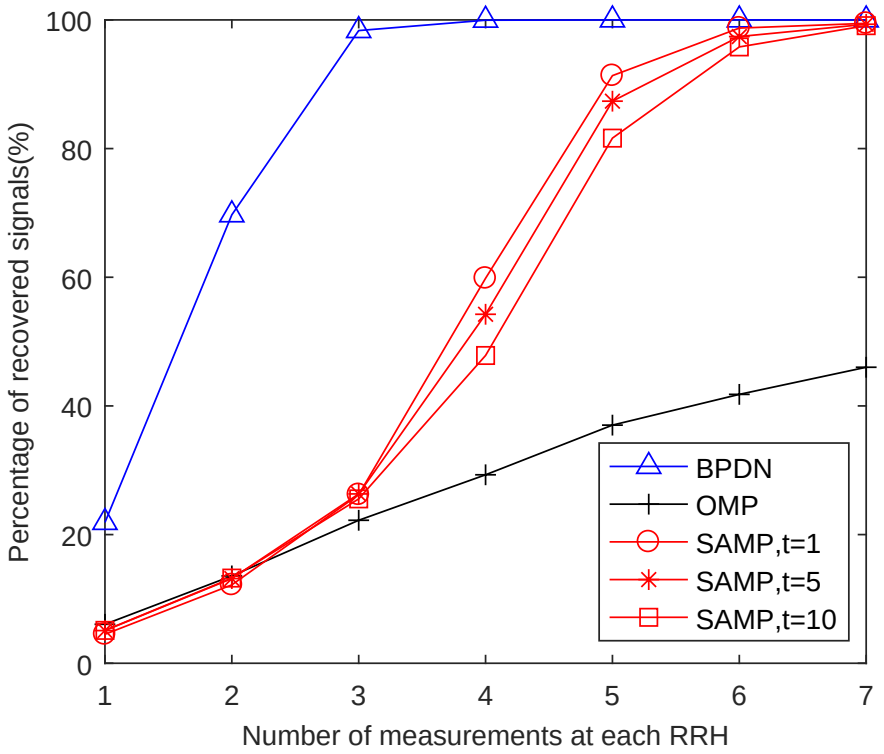
In the following sections, we will investigate the performance of two proposed schemes by using this indicator $P_{\text{recovered}}$.

In addition, an OMP-based recovery scheme will also be considered as a performance benchmark. To be specific, OMP is used to detect the set of active users instead of BPDN and SAMP, and then the ZF receiver is used to recover the original signals.

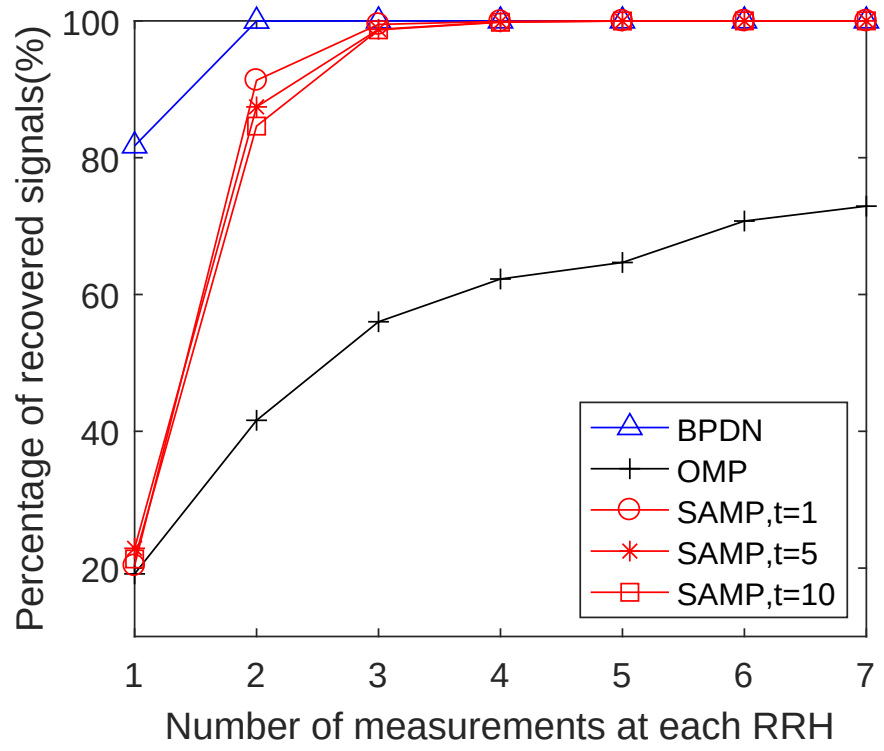
4.5.2 The Percentage of Exact Recovered Signals Versus Number of Measurements

Set $s = 20$ and choose $M = 20, 50$ for comparison. Fig. 4.4 compares the recovery performance of BPDN, SAMP and OMP under different scenarios. It is evident that the

percentage of recovered signals with the SAMP algorithm decreases dramatically when only a small number of RRHs is deployed. SAMP requires at least $R = 6$ measurements at each RRH to achieve recovery of roughly 95% of source signals, whereas BPDN can reach the same performance with only half the required measurements ($R = 3$) at each RRH. BPDN, however, still cannot guarantee exact recovery of all signals, especially when the total number of measurements is less than four times the sparsity. This numerical result confirms the so-called four-to-one practical rule (i.e., for exact recovery, the ratio of number of measurements to sparsity need to be about $4 : 1$) [30]. In contrast, given a larger number of RRHs, both BPDN and SAMP will have better recovery performances. The percentage of accurately recovered signal quickly approaches 100% as the number of measurements increases. However, the OMP fails to recover all the signals even with $R = 7$ measurements at each RRH. This is beacause the OMP can not detect the whole set of active users accurately even provided the sparsity as a *prior* information.



(a) $M = 20$



(b) $M = 50$

Figure 4.4: Percentage of recovered signals versus number of measurements under number of active users and the number of active users is $s = 20$. (a) $M = 20$. (b) $M = 50$.

Fig. 4.5 shows the recovery performance of BPDN with different numbers of RRHs deployed. It is evident that when the number of measurements at each RRH $R < 3$, the percentage of recovered signals increases significantly with increasing numbers of RRHs. In other words, for exact recovery using BPDN, the number of measurements at each RRH should be properly selected according to the deployment of RRHs. This is again consistent with statement of four-to-one rule [30]. Fig. 4.6 shows the performance of accurately recovering signals using SAMP algorithm (step size $t = 5$) with the number of RRHs $M = 20, 30, 40, 50, 60$. Similar to the BPDN recovery scheme, the percentage of recovered signals increases as the number of RRHs increases.

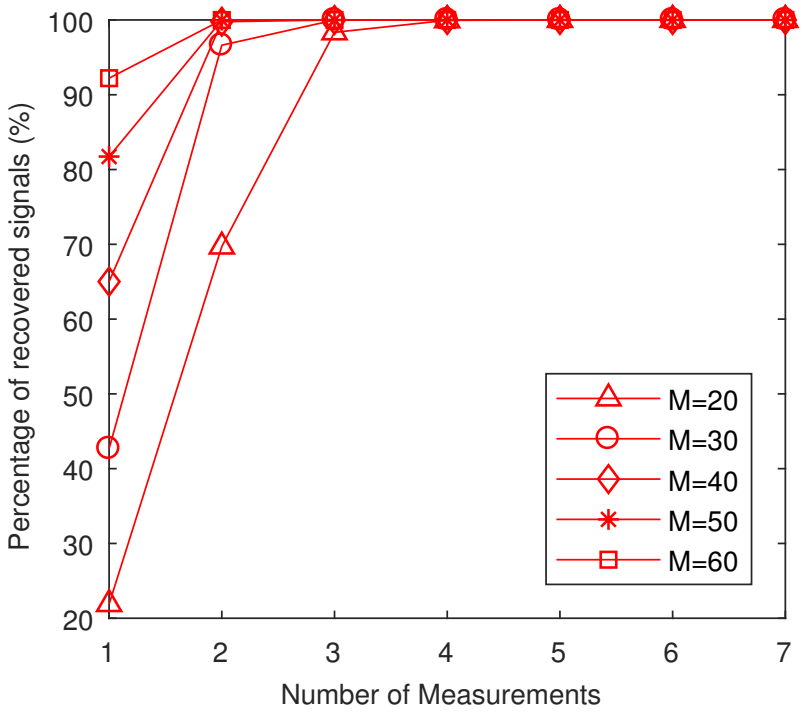


Figure 4.5: Percentage of recovered signals versus number of measurements at each RRH via BPDN, for various numbers of RRHs $M = 20, 30, 40, 50, 60$ and $s = 20$ active users.

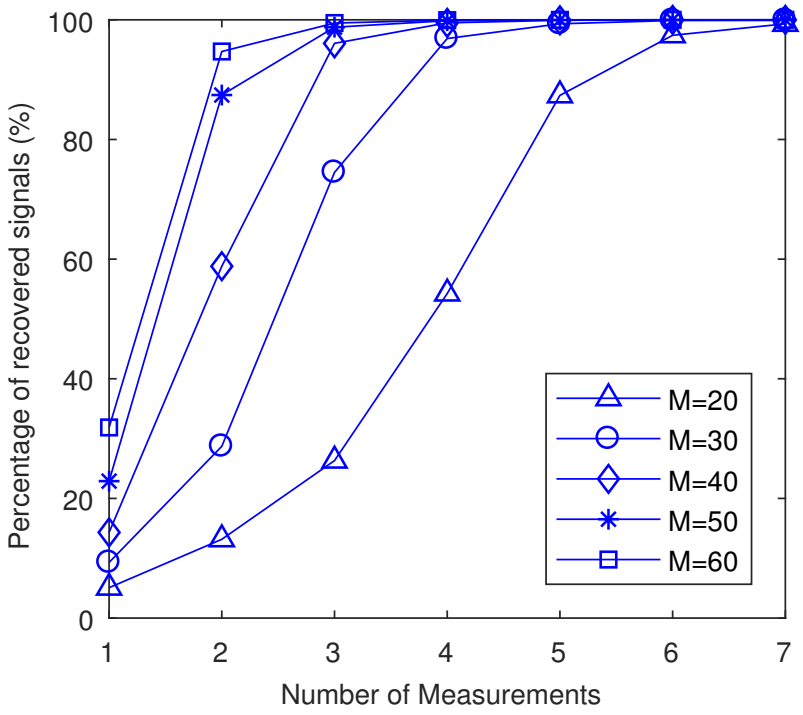


Figure 4.6: Percentage of recovered signals versus number of measurements at each RRH via SAMP (with step size $t = 5$), for various numbers of RRHs $M = 20, 30, 40, 50, 60$ and $s = 20$ active users.

From these figures, we can conclude that with a proper number of RRHs deployed, the proposed scheme, using either the BPDN or SAMP recovery algorithm, can guarantee exact recovery given a sufficient number of measurements. SAMP has lower complexity but shows worse performance compared to BPDN. This is therefore a trade-off between performance and computational complexity. On the other hand, the proposed scheme can accurately recover the signals using many fewer measurements than the conventional ZF receiver which usually requires more antennas (and thus more measurements) than the total number of users, and hence the fronthaul load will be effectively reduced.

4.5.3 The Percentage of Exact Recovered Signals Versus Sparsity

Fig. 4.7 illustrates the percentage of successfully recovered signals versus the number of active users, given a fixed number of measurements $R = 4$ at each RRH. It can be seen that for a binary sparse signal, the BPDN scheme has much better performance as the

sparsity level increases. However, SAMP starts to fail to recover all active users when the number of active users $s > 50$. This is because the number of measurements becomes insufficient for SAMP to complete exact recovery when the the signal tends to be less sparse. Again, the OMP fails to recover all the signals due to the inaccurate detection of the set of active users.

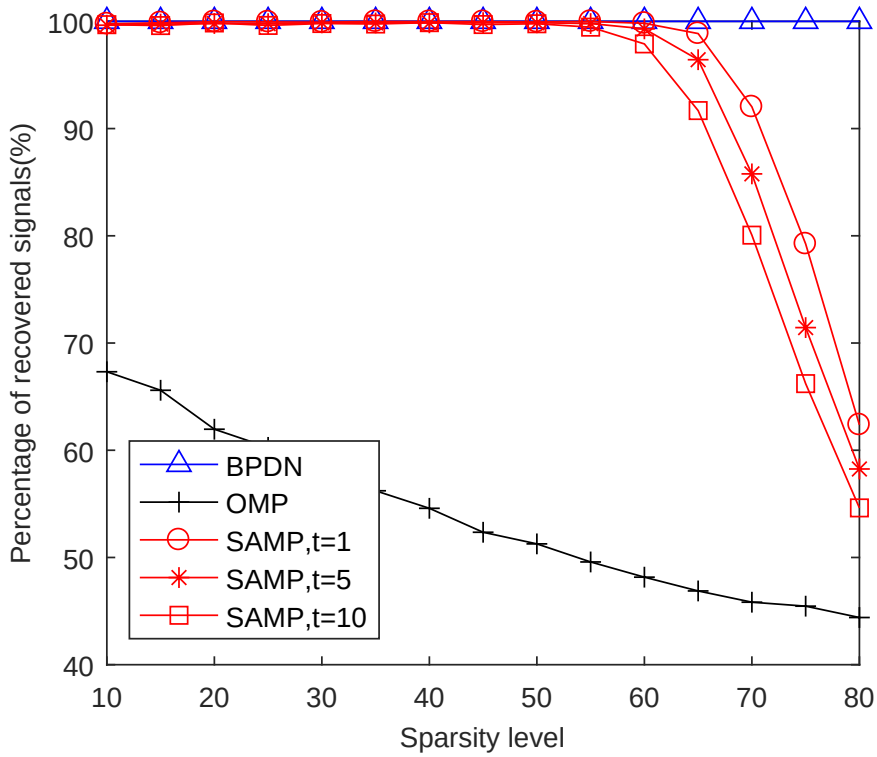


Figure 4.7: Percentage of recovered signals versus number of active users. Here, the number of RRHs $M = 50$ and the number of measurements at each RRH $R = 4$.

4.6 Summary

A compression and recovery scheme based on compressive sensing has been proposed for the uplink of a cell-free C-RAN system in this chapter. The whole scheme is divided into three stages: 1) Signal reception at RRHs; 2) Fronthaul compression and forwarding to the BBU pool; 3) Signal recovery at the BBU pool. In the signal recovery stage, the BPDN and SAMP algorithms are adopted respectively to realise transmitted signal estimation. The numerical results show that the proposed scheme can achieve effective signal recovery using either BPDN or SAMP recovery algorithm and relieve the burden

on the fronthaul link.

Chapter 5

Conclusions and Further Work

Contents

5.1 Conclusions 57

5.2 Further Work 58

5.1 Conclusions

In this thesis, we have focused on fronthaul load reduction in the uplink of a cell-free C-RAN system. A compressive sensing based compression and recovery scheme has been proposed for handling this issue.

In Chapter 3, the basic concept of compressive sensing has been introduced. First, a standard compressive sensing model along with the essential notions of sparsity, incoherence has been presented. Secondly, the robustness of compressive sensing has been discussed by introducing a practical compressive sensing model and a tool to study the robustness, called the restricted isometry property (RIP). In particular, we have provided some potential measurement matrices which have been verified to satisfy the RIP condition and therefore facilitate signal recovery with high probability. Thirdly, a variety of popular recovery algorithms in compressive sensing has been illustrated in detail, including basis pursuit denoising (BPDN), orthogonal matching pursuit (OMP), compressive sampling matching pursuit (CoSaMP) and sparsity adaptive matching pursuit (SAMP).

Furthermore, we have provided a comparison among these recovery algorithms, mainly in terms of the computational complexity and the minimal number of measurements required for reconstructing a sparse signal. The study of compressive sensing was presented as a preliminary for the following study on fronthaul compression and signal recovery strategy for an uplink C-RAN system.

In Chapter 4, a compression and recovery scheme based on the compressive sensing technique has been proposed for an uplink cell-free C-RAN system, where a very large number of multi-antenna RRHs simultaneously serve a smaller number of single-antenna users over a service area, in which there are no cells or cell boundaries and therefore is described as "cell-free". The overall scheme was accomplished by three stages: 1) signal gathering at each RRH, 2) fronthaul compression and forward, and 3) signal recovery at the BBU pool. More precisely, a conventional compressive sensing problem which incorporates the multi-access channel fading has been formulated for reducing the fronthaul load as well as recovering the transmitted signals from users. Both the BPDN and SAMP algorithms have been applied to estimate the sparse signals, and the analysis of the stability of BPDN and SAMP recovery has been given. Finally, the performance of proposed scheme has been presented by the simulation results. It has been shown that our proposed scheme achieves robust signal recovery and efficient fronthaul load reduction.

5.2 Further Work

Based on the content of this thesis, some possible directions for further work are given below:

- As the proof of the RIP for the measurement matrix used in this thesis refers to a great deal of mathematics which is beyond the scope of the present work, we only provide some theorems that are proved by the previous work of other researchers. The proof in our case will therefore be an interest for future study.
- The recovery algorithms employed in our proposed scheme are classic approaches: BPDN is shown to accurately recover the sparse signal with fewer measurements but its computational complexity will be considerably high; the SAMP algorithm

can offer lower computational complexity but at the cost of recovery accuracy. So it is worth finding an improved recovery algorithm that can reach a compromise between recovery accuracy and computational complexity, meanwhile, it should not require the sparsity of the signal as prior information.

- The system in this thesis considers only a single-input multiple-output (SIMO) scenario, which can be extended to a multiple-input multiple-output (MIMO) scenario where both the RRHs and users are equipped with multiple antennas. In this case, the coherence among the received signals at different antennas within each RRH can be exploited for the sparse representation of received signals. On the other hand, interference mitigation will be another issue needing to be solved.

Glossary

BBU	**B**ase**B**and **U**nit
BPDN	**B**asis **P**ursuit **D**e**N**oising
BS	**B**ase **S**tation
CAP	**C**ompression-**A**fter-**P**recoding
CAPEX	**CAP**ital **EX**penditure
CBP	**C**ompression-**B**efore-**P**recoding
CCD	**C**harged-**C**oupled **D**evice
CMOS	**C**omplementary **M**etal-**O**xide-**S**emiconductor
CoMP	**Co**operative **M**ulti-point **P**rocessing
CoSaMP	**Co**mpressive **Sa**mpling **M**atching **P**ursuit
C-RAN	**C**loud **R**adio **A**ccess **N**etwork
CSI	**C**hannel **S**tate **I**nformation
eICIC	**e**nhanced **I**nter-**C**ell **I**nterference **C**oordination
FD	**F**ull **D**uplex
F-RAN	**F**og-computing-based **R**adio **A**ccess **N**etwork
H-CRAN	**H**eterogeneous **C**loud **R**adio **A**ccess **N**etwork
HetNets	**Het**erogeneous **Net**work**s**
ICI	**I**nter-**C**ell **I**nterference
I/Q	**I**n-phase and **Q**uadrature
JT	**J**oint **T**ransmission
MAC	**M**edium **A**ccess **C**ontrol
MIMO	**M**ultiple-**I**nput **M**ultiple-**O**utput
mMTC	**m**assive **M**achine-**T**ype **C**ommunication
MRI	**M**agnetic **R**esonance **I**maging

OMP	**O**thogonal **M**atching **P**ursuit
OPEX	**OP**erating **EX**penditure
PHY	**PHY**sical layer
RAN	**R**adio **A**ccess **N**etwork
RANaaS	**R**adio **A**ccess **N**etwork-**a**s-**a**-**S**ervice
RF	**R**adio **F**requency
RIP	**R**estricted **I**sometry **P**roperty
RoF	**R**adio-**o**ver-**F**ibre
ROMP	**R**egularized **O**rthogonal **M**atching **P**ursuit
RRH	**R**emote **R**adio **H**ead
SAMP	**S**parsity **A**daptive **M**atching **P**ursuit
SIC	**S**uccessive **I**nterference **C**ancellation
SISO	**S**ingle-**I**nput **S**ingle-**O**utput
SP	**S**ubspace **P**ursuit
SQNR	**S**ignal-to-**Q**uantization-**N**oise-**R**atio
StOMP	**St**agewise **O**rthogonal **M**atching **P**ursuit
UUP	**U**niform **U**ncertainty **P**rinciple
WDM	**W**avelength **D**ivision **M**ultiplexing
WNC	**W**ireless **N**etwork **C**loud
ZF	**Z**ero **F**orcing

Bibliography

[1] C. Mobile, "C-RAN: The road towards green RAN," *White Paper*, 2011.

[2] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Pape," Mar. 2017 [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[3] Y. Lin, L. Shao, Z. Zhu, Q.Wang, and R. K. Sabhikhi, "Wireless network cloud: Architecture and system requirements," *IBM J. Res. Develop.*, vol.54, no.1, pp.4:1-4:2, Jan./Feb. 2010.

[4] A. Checko *et al.*, "Cloud-RAN for mobile networks—A technology overview," *IEEE Commun. Surveys Tuts.*, vol.17, no.1, pp.405-426, Sep. 2014.

[5] G. Kardaras and C. Lanzani, "Advanced multimode radio for wireless and mobile broadband communication," *Proc. EuWIT Conf.*, pp.132-135.

[6] "ZTE green technology innovations white paper," Shenzhen, China, 2011, Tech. Rep.

[7] "Intel heterogenous network solution brief," Santa Clara, CA, USA, 2011, Tech. Rep.

[8] B. Haberland *et al.*, "Radio base stations in the cloud," *Bell Labs Tech. J.*, vol.18, no.1, pp.129-152, 2013.

[9] D. Sabella et al., "RAN as a service: Challenges of designing a flexible RAN architecture in a cloud-based heterogeneous mobile network," in *Proc. Future Netw. Summit*, Jul. 2013, pp.1-8.

[10] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): A primer," *IEEE Netw.*, vol.29, no.1, pp.35-41, Jan. 2015.

[11] C. Liu, L. Zhang, M. Zhu, J. Wang, L. Cheng and G. Chang, "A Novel Multi-Service Small-Cell Cloud Radio Access Network for Mobile Backhaul and Computing Based on Radio-Over-Fiber Technologies," *J. Lightw. Technol.*, vol.31, no.17, pp.2869-2875, Sep. 1, 2013.

[12] M. Peng, Y. Li, J. Jiang, J. Li, and C.Wang, "Heterogeneous cloud radio access networks: A new perspective for enhancing spectral and energy efficiencies," *IEEE Wireless Commun.*, vol.21, no.6, pp.126-135, Dec. 2014.

[13] M. Peng, S. Yan, K. Zhang, and C. Wang. "Fog Computing Based Radio Access Networks: Issues and Challenges," *IEEE Net.*, vol.30, no.4, pp.46-53, Jul. 2016.

[14] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Commun. Surveys Tuts.*, vol.18, no.3, pp.2282-2308, Aug. 2016.

[15] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Commun.*, vol.22, no.2, pp.152-160, Apr. 2015.

[16] S. Park, O. Simeone, O. Sahin, and S. S. Shitz, "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Process. Mag.*, vol.31, no.6, pp.69-79, Oct. 2014.

[17] S. Park et al., "Robust and efficient distributed compression for cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol.62, no.2, pp.692-703, Feb. 2013.

[18] Y. Zhou and W. Yu, "Optimized backhaul compression for uplink cloud radio access network," *IEEE J. Sel.Areas Commun.*, vol.32, no.6, pp.1295-1307, Jun. 2014.

[19] Y. Zhou and W. Yu, "Fronthaul compression and transmit beamforming optimization for multi-antenna uplink C-RAN," *IEEE Trans. Signal Process.*, vol.64, no.16, pp.4138-4151, Aug. 2016.

[20] X. Rao and V. Lau, "Distributed fronthaul compression and joint signal recovery in cloud-RAN," *IEEE Trans. Signal Process.*, vol.63, no.4, pp.1056-1065, Feb. 2015.

[21] Y. Wang, Z. Chen and M. Shen, "Compressive sensing for uplink cloud radio access network with limited backhaul capacity," in *4th ICCSNT*, 2015, pp.898-902.

[22] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Process.*, vol.61, no.22, pp.5646-5658, Nov. 2013.

[23] S. Park, O. Simeone, O. Sahin, and S. Shamai, "Inter-cluster design of precoding and fronthaul compression for cloud radio access networks," *IEEE Wireless Commun. Lett.*, vol.3, no.4, pp.369-372, Apr. 2014.

[24] K. Nguyen, Q. Vu, M. Juntti and L. Tran, "Energy efficient precoding C-RAN downlink with compression at fronthaul," *IEEE Int. Conf. Commun.*, May 2017, pp.1-6.

[25] J. Kang, O. Simeone, J. Kang, and S. Shamai, "Fronthaul compressionand precoding design for C-RANs over ergodic fading channels," *IEEE Trans. Veh. Technol.*, vol.65, no.7, pp.5022-5032, Jul. 2016.

[26] Y. Jeon, S. Park, C. Song, J. Moon, S. Maeng and I. Lee, "Joint designs of fronthaul compression and precoding for full-duplex cloud radio access networks," *IEEE Wireless Commun. Lett.*, vol.5, no.6, pp.632-635, Dec. 2016.

[27] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol.52, no.2, pp.489-509, Feb. 2006.

[28] E. Candes, " Compressive sampling," Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.

[29] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol.52, no.4, pp.1289-1306, Apr. 2006.

[30] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol.25, no.2, pp.21-30, Mar. 2008.

[31] M. Duarte, M. Davenport, D. Takhar, J. Laska, S. Ting, K. Kelly, R.G. Baraniuk, "Single-Pixel Imaging via Compressive Sampling," *IEEE Signal Process. Mag.*, vol.25, no.4, pp.83-91, Mar. 2008.

[32] J. Haldar, D. Hernando, Z. Liang, "Compressed-sensing MRI with random encoding," *IEEE Trans. Med. Imag.*, vol.30, no.4, pp.893-903, Oct. 2011.

[33] E. candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol.51, no.12, pp.4203-4215, Dec. 2005.

[34] W. Bajwa, J. Haupt, A.M. Sayeed, R. Nowak, "Compressed channel sensing: a new approach to estimating sparse multipath channels," *Proc. IEEE*, vol.98, no.6, pp.1058-1076, Jun. 2010.

[35] S. Foucart and H. Rauhut, *An Mathematical Introduction to Compressive Sensing*. New York, NY, USA: Springer-Verlag, 2010.

[36] D. Donoho, M. Elad, "Optimally sparse representations in general (non-orthogonal) dictionaries via l_1 minimization," *Proc. Nat. Acad. Sci.*, vol.100, no.5, pp.2197-2202, Mar. 2003.

[37] E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol.52, no.12, pp.5406-5425, Dec. 2006.

[38] W. Bajwa, J. Haupt, G. Raz, S. Wright, R. Nowak, "Toeplitz-structured compressed sensing matrices," In *Proc. IEEE Stat. Sig. Proc. Workshop*, pp.294-298, 2007.

[39] R. DeVore. "Deterministic constructions of compressed sensing matrices," *Journal of complexity*, vol.23, pp.918-925, May 2007.

[40] M. Elad, *Spare and Redundant Representations*. New York, NY, USA: Springer-Verlag, 2013.

[41] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[42] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol.59, no.8, pp.1207-1223, Aug. 2006.

[43] I. E. Nesterov, A. Nemirovskii, and Y. Nesterov, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.

[44] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol.53, no.12, pp.4655-4666, Dec. 2007.

[45] A. Bjorck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA: SIAM, 1996.

[46] S. Kunis and H. Rauhut, "Random sampling of sparse trigonometric polynomials II: Orthogonal matching pursuit versus basis pursuit," *Found. Comp. Math.*, vol.8, no.6, pp.737-763, Dec. 2008.

[47] D. Needell, J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol.26, no.3, pp.301-321, July 2008.

[48] D. Donoho, Y. Tsaig, I. Drori and J. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol.58, no.2, pp.1094-1121, Feb. 2012.

[49] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol.4, no.2, pp.310-316, Apr. 2010.

[50] W. Dai, O. Milenkovic, "Subspace Pursuit for Compressive Sensing Signal Reconstruction," *IEEE Trans. Inform. Theor.*, vol.55, no.5, pp.2230-2249, May 2009.

[51] T. Do, L. Gan, N. Nguyen and T. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proc. 42th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, 2008, pp.581-587.

[52] H. Ngo, A. Ashikhmin, H.Yang, E. Larsson and T. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol.16, no.3, pp.1834-1850, Mar. 2017.

[53] X. Wang, M. Sheng, Y. Lou, Y. Shih and M. Chiang, "Internet of things session management over LTE balancing signal load, power, and delay," *IEEE Internet Things J.*, vol.3, no.3, pp.339-353, June 2016.

[54] 3GPP TS 22.368, "Service Requirements for Machine-Type Communications (MTC)," V14.0.1, Oct. 2017.

[55] Y. Jiang, M. Varanasi and J. Li, “Performance analysis of ZF and MMSE equalizers for MIMO systems: An in-depth study of the high SNR regime,” *IEEE Trans. Inf. Theory*, vol.57, no.4, pp.2008-2026, Mar. 2011.