Compressive Sensing Based Grant-Free Communication

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy by

Yuanchen Wang

Department of Electrical Engineering and Electronics
School of Electrical Engineering, Electronics and Computer Science

May 2022

Abstract

Grant-free communication, where each user can transmit data without following the strict access grant process, is a promising technique to reduce latency and support massive users. In this thesis, compressive sensing (CS), which exploits signal sparsity to recover data from a small sample, is investigated for user activity detection (UAD), channel estimation, and signal detection in grant-free communication, in order to extract information from the signals received by base station (BS).

First, CS aided UAD is investigated by utilizing the property of quasi-time-invariant channel tap delays as the prior information for the burst users in internet of things (IoT). Two UAD algorithms are proposed, which are referred to as gradient based and time-invariant channel tap delays assisted CS (g-TIDCS) and mean value based and TIDCS (m-TIDCS), respectively. In particular, g-TIDCS and m-TIDCS do not require any prior knowledge of the number of active users like the existing approaches and therefore are more practical.

Second, periodic communication as one of the salient features of IoT is considered. Two schemes, namely periodic block orthogonal matching pursuit (PBOMP) and periodic block sparse Bayesian learning (PBSBL), are proposed to exploit the non-continuous temporal correlation of the received signal for joint UAD, channel estimation, and signal detection. The theoretical analysis and simulation results show that the PBOMP and PBSBL outperform the existing schemes in terms of the success rate of UAD, bit error rate (BER),

and accuracy in period estimation and channel estimation.

Third, UAD and channel estimation for grant-free communication in the presence of massive users that are actively connected to the BS is studied. An iteratively UAD and signal detection approach for the burst users is proposed, where the interference of the connected users on the burst users is reduced by applying a preconditioning matrix to the received signals at the BS. The proposed approach is capable of providing significant performance gains over the existing algorithms in terms of the success of UAD and BER.

Last but not least, since the physical layer security becomes a critical issue for grant-free communication, the channel reciprocity in time-division duplex systems is utilized to design environment-aware (EA) pilots derived from transmission channels to prevent eavesdroppers from acquiring users' channel information. The proposed EA-pilots based approach possesses a high level of security by scrambling the eavesdropper's normalized mean square error performance of channel estimation.

Keywords: Grant-free communication, compressive sensing, user activity detection, channel estimation, signal detection.

Acknowledgements

Thinking back on the past four years of my Ph.D. life, I could not have come through without the guidance and support from all the great people around me. First and most importantly, I would like to give my most sincere gratitude to my supervisors, Prof. Eng Gee Lim and Prof. Xu Zhu. They have been incredible mentors to me in academic study and attitude towards life. Prof. Eng Gee Lim and Prof. Xu Zhu have always delivered well-designed, substantial content in a casual atmosphere. It helped me well throughout my Ph.D. study. My gratitude also goes to my supervisor, Prof. Yi Huang, for helping me deal with student affairs at University of Liverpool.

Second, I would like to give my sincere appreciation to Dr. Zhongxiang Wei, Dr. Yujie Liu and Dr. Lin Gan. They have provided me with countless valuable advices. Dr. Zhongxiang Wei has offered me tremendous help with valuable guidance in research, academic writing, and support regarding the equipment.

Third, I would like to give special thanks to Dr. Yufei Jiang, Mr. Haiyong Zeng, Mr. Jie Cao, Mr. Hanjun Duan, and my friends from Harbin Institute of Technology, Shenzhen, for their valuable insights in both group meetings and my annual reviews. My life in Shenzhen was productive and enjoyable.

I would also like to thank my roommate, Dr. Rui Pei, and my friends from EE414 and EE511. All these friends made my Ph.D. journey enjoyable.

I would like to save special thanks to my beloved family for their loving consideration and great confidence in me all through these years. Furthermore, I would like to dedicate this thesis to my son, Yishen (Ethan) Wang. It is him who brings me hope and pleasure when I hesitate to move forward in my research.

Last but not least, the department of Electrical Engineering and Electronics of the University of Liverpool and the department of Electrical and Electronic Engineering of Xi'an Jiaotong-Liverpool University, the Electronic and Information Engineering Department, Harbin Institute of Technology, Shenzhen are gratefully acknowledged for their support.

Contents

List of Figures

List of Tables

List of Abbreviation

3GPP	The Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AoA	Angle of Arrival
AP	Access Point
AWGN	Additive White Gaussian Noise
BCoSaMP	Block Compressive Sampling Matching Pursuit
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
BSR	Buffer Status Report
CDMA	Code Division Multiple Access
CI	Constructive Interference
CP	Cyclic Prefix

CRLB	Cramér-Rao Lower Bound
CS	Compressive Sensing
CSI	Channel State Information
DACS	Dynamic Adaptive CS
DCS	Dynamic CS
DFT	Discrete Fourier Transform
DPC	Dirty-Paper Coding
E2E	End-to-End
EA	Environment-Aware
EMP	Efficient Matching Pursuit
Eve	Eavesdropper
FFT	Fast Fourier Transform
FIM	Fisher Information Matrix
g-TIDCS	Time-Invariant Channel Tap Delays Assisted CS
GSM	Global System for Mobile Communications
ICASBL	Independent Component Analysis Sparse Bayesian Learning
IE	Interference Utilization
IFFT	Inverse FFT
IM	Interference Mitigation
IORLS	Iterative Order Recursive Least Square

IoT	Internet of Things
IS	Interference Suppression
ITU	International Telecommunication Union
LDS-CDMA	Low-Density Spread Spectrum CDMA
LDS-OFDM	Low-Density Spread Spectrum based OFDM
LMMSE	Linear Minimum Mean Squared Error
LS	Least Square
LTE	Long Term Evolution
LTE-A	LTE Advanced
m-TIDCS	Mean Value based and TIDCS
MA	Multiple Access
MAC	Medium Access Control
MIMO	Multiple-Input Multiple-Output
mMTC	Massive MTC
MP-BSBL	Message Passing based Block SBL
MTC	Machine Type Communication
MTCD	MTC Device
NMSE	Normalized Mean Square Error
NOMA	Non-Orthogonal MA
OFDM	Orthogonal Frequency Division Multiplexing

OFDMA Frequency Division Multiple Access

OMA Orthogonal MA

OMP Orthogonal Matching Pursuit

PBOMP Periodic Block Orthogonal Matching Pursuit

PBSBL Periodic Block Sparse Bayesian Learning

PDCCH Physical Downlink Control Channel

PDMA Pattern Division Multiple Access

PHY Physical Layer Security

PIA-ASP Prior-Information Aided Adaptive Subspace Pursuit

PRACH Physical Random Access Channel

PUCCH The Physical Uplink Control Channel

QoS Quality of Service

RA Random Access

RER Relatively Error Ratio

RIP Restricted Isometry Property

RRC Radio Resource Control

SBL Sparse Bayesian Learning

SCMA Sparse Code Multiple Access

SDMA Spatial Division Multiple Access

SIC Successive Interference Cancellation

SINR	Signal to Interference Plus Noise Ratio
SNR	Signal-to-Noise Ratio
SR	Scheduling Request
TDD	Time Division Duplex
THP	Tomlinson-Harashima Precoding
TTI	Transmission Time Interval
UAD	User Activity Detection
UE	User Equipment
ULA	Uniform Linear Array
URLLC	Ultra-Reliable and Low Latency Communications
VP	Vector Perturbation
ZF	Zero-Forcing

List of Symbols

Notations

$(\cdot)_\Gamma$	Entries of a vector in the set Γ or submatrix comprising the Γ columns of a matrix, depending on the context
$(\cdot)_\mathrm{I}$	Imaginary part of the complex argument
$(\cdot)^{-1}$	Matrix inverse
$(\cdot)^\dagger$	Pseudo inverse
$(\cdot)^\mathrm{H}$	Conjugate transpose
$(\cdot)^\top$	Vector transpose
$\Gamma \backslash \widehat{\Gamma}$	Set consisting of elements in Γ while not in $\widehat{\Gamma}$
$\lfloor \cdot \rfloor$	Operation of round down
$\mathbb{E}_x$	Expectation with respect to the random variable x
$\mathbf{F}_L$	Discrete fourier transform matrix with the first L columns
$\mathbf{I}$	Identity matrix

$\mathcal{CN}(\mu, \sigma^2)$ Circularly symmetric complex Gaussian distribution with mean μ and σ^2 variance

$\mathcal{C}$ Field of complex numbers

$\otimes$ Kronecker product

$\Re$, $(\cdot)_{\mathrm{R}}$ Real part of the complex argument

$\mathrm{Cov}(\cdot)$ Covariance

$\mathrm{diag}(\cdot)$ Diagonal matrix whose diagonal elements are a vector

$\exp(\cdot)$ Exponent function with base of e

$\mathrm{Tr}(\cdot)$ Pseudo inverse

$\mathrm{var}(x)$ Variance with respect to the random variable x

$\mathrm{vec}(\cdot)$ Column-ordered vectorization of matrix

$\mathrm{vec}(\cdot)^{-1}$ Inverse process of vectorization

$p(x)$ Probability density function of the random variable x

Boldface lower case letters Vectors

Boldface upper case letters Matrices

Parameters

J Number of symbols

K Number of active users

L Number of path

L_M Maximum path delay

M Number of subcarriers or number of connected users

N Total number of potential active users

P Transmission period

Q Block length

R Number of antennas of base station

S Spreading sequence

V Total number of users

Chapter 1

Introduction

1.1 Background

Future mobile communications is focused on providing seamless coverage for the massive number of machines and devices building the Internet of Things (IoT), as well as personal communications [1, 2, 3, 4, 5, 6]. New applications such as tele-medicine, tele-surgery, and intelligent transportation present new specifications for reliability, throughput, and end-to-end (E2E) latency [7, 8]. Hence, large amounts of data are collected by aperiodic and/or periodic communication [9, 10]. In particular, there are massive devices are fixed in a place without moving or in a low-speed moving scenario. For example, control center, factory, warehouse and logistics form a complex communication environment in manufacturing, where the massive devices, *e.g.,* sensors and actuators, periodically and/or aperiodically report their pressure, humidity, or temperature parameters to the base station (BS).

Figure 1.1 depicts various wireless connectivity options that support different use cases in the IoT framework [11]. A basic classification of these technologies for different IoT use cases is made by considering network coverage and performance criteria. It is expected that a large portion of these devices will be implemented over a wide area network enabled mainly by cellular networks. In this context, the third generation partnership project (3GPP) provides connectivity over cellular networks, including Global System for Mobile Communications (GSM) [12], Code Division Multiple Access (CDMA) [13], orthogonal frequency division multiplexing (OFDM) based Long Term Evolution (LTE) [14], and the upcoming the fifth generation (5G) communications [15]. These technologies operate on a licensed spectrum and are evolving for low-power IoT applications.

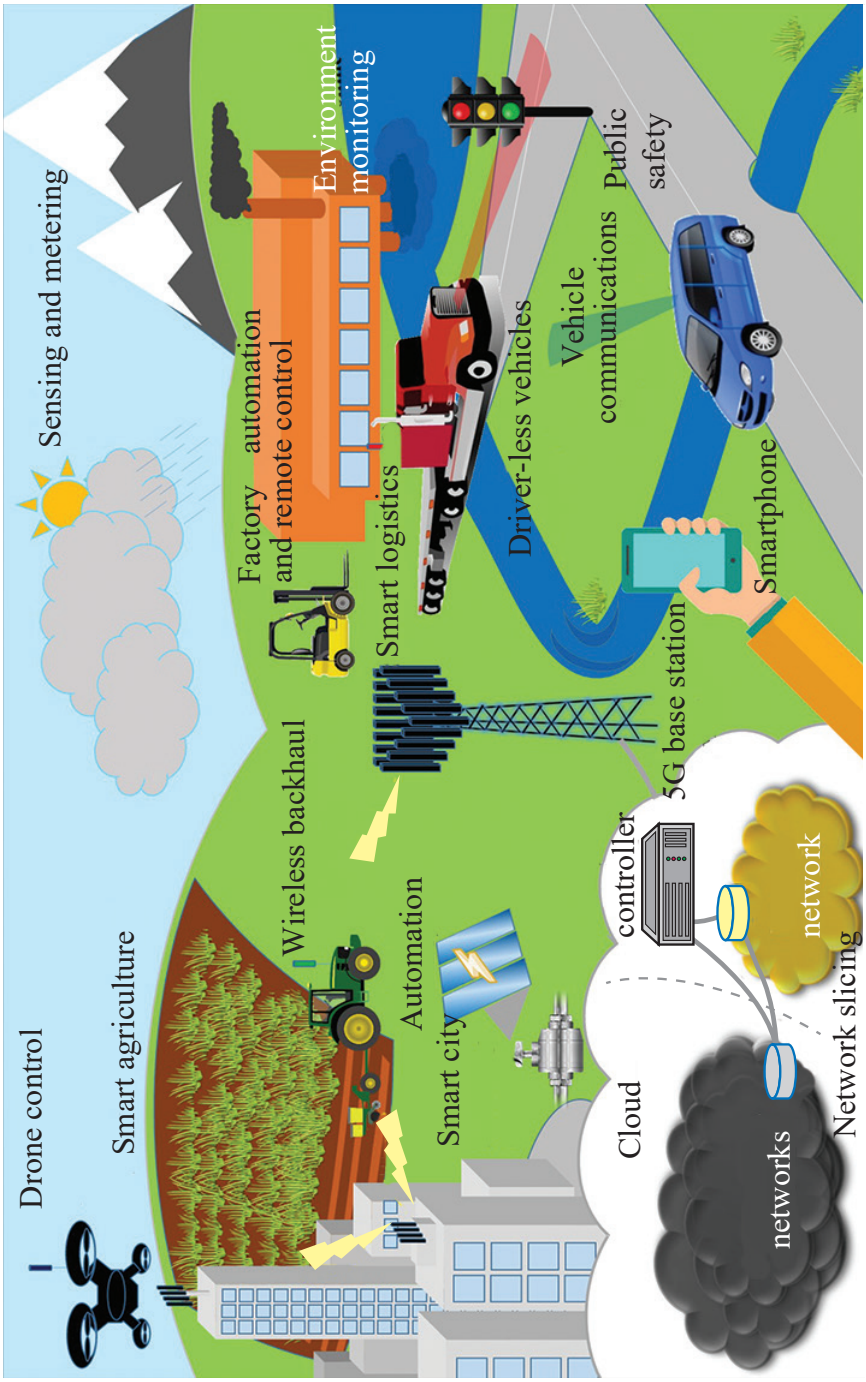

Figure 1.1: Overview of communication systems for IoT

1.1.1 Grant-Free Communication

Several emerging technologies, including wearables, virtual/augmented reality, and fully immersive experiences, are shaping human end-users' demeanor with their specific requirements. Hence, these use cases for next generation networks are driving specifications for multiple aspects of data rate, latency, reliability, device/network energy efficiency, traffic density, mobility, and connection density. Current fourth generation (4G) or 5G networks cannot meet all the technical requirements for these services [16]. The future communication is a wireless access solution that meets the specifications for wireless communications in 2025 and beyond. Based on consensus of International Telecommunication Union (ITU), future communications provide machine type communication (MTC), personal mobility services and latency/reliability critical services. Particularly, E2E latency down to 1 ms needs to be met for future communications [17, 18].

Latency is critical in many applications such as transportation, automated industrial production, healthcare, robotics/control, entertainment, education, virtual reality, and culture. IoT is rapidly becoming a reality that connects anything, anywhere and anytime. Currently, the existing IoT applications are provided by LTE/LTE Advanced (LTE-A), however, some applications have more stringent requirements on the underlying network, such as low latency, high reliability, and security [19].

In general, E2E latency can be divided into two main components: access latency and transmission latency. Delays in packet transmission can be caused by the radio access network, backhaul, core network, and data center/Internet [20]. The access delay is defined as the transition time for the user equipment (UE) to switch from the idle state to the active state. The UE is not connected to the Radio Resource Control (RRC) in the idle state. After the RRC connection is established, the UE switches from the idle state to the connected state. However, due to the large number of conflicts in massive IoT scenarios, many users still cannot access the network. As a result, the access latency may vary significantly compared to the packet transmission latency T. Building upon this structure, access latency is considered as a performance bottleneck for future wireless communication.

Radio resources, *e.g.*, time and frequency, are orthogonally allocated to connected devices in existing wireless networks. Therefore, in LTE/LTE-A, users requesting access to the cellular network must first align with BS through a contention-based process on the physical random access channel (PRACH), which (in LTE/LTE-A) has been identified as a significant performance bottleneck, resulting in excessive latency and signaling overhead.

The process was initially intended to be used for multiple purposes, *i.e.,* initial access for users not connected to the BS, uplink synchronization for connected users, data transfer or acknowledgment response, handover management, *etc.* The LTE shows that PRACH access is performed by a contention-based four-way handshake (the user can initiate the access process at any time) [21].

The device is first notified of available PRACH resources via a broadcast from the BS to enable PRACH access, followed by a four-way handshake. Any device already aligned with the BS can release an access request to transmit its data. The random access (RA) of LTE/LTE-A is shown in Figure 1.2. In the current LTE system, the UE sends a scheduling request (SR) and a buffer status report (BSR) to request uplink authorization from the BS. The UE need to wait for an scheduling request (SR) opportunity on the physical uplink control channel (PUCCH) resource to send an SR. Upon receiving the SR, the BS sends an uplink authorization to the corresponding UE on the physical downlink control channel (PDCCH)). Then, the signals generated by UE is transmitted over the physical uplink shared channel (PUSCH) on the scheduling resource.

There are many problems associated with RA process illustrated by Figure 1.2, especial for the users that sporadically wake up from idle state, referred to as burst users. If there is PRACH congestion due to a large number of MTCs trying to access at the same time, the signal-to-noise ratio (SNR) observed at the BS may be very low, where messages cannot be detected, and therefore, many access fail. Excessive preamble collisions and retransmissions lead to network congestion, packet loss, unexpected delays, high energy consumption, and wasted radio resources. In contrast, the idea of grant-free communication is that each burst user transmits its data directly to the BS without waiting for any permission. Therefore, it is perfect for latency-sensitive IoT scenarios.

Table. 1.1 provides detailed information on the comparison of latency between grant-based and grant-free scheduling in terms of transmission time interval (TTI) [21]. For 1 TTI = 1 orthogonal frequency division multiple access (OFDMA) symbol, the delays for grant-free and grant-based transmissions can be calculated as $4 \times 71.354\ ms = 0.285\ s$ and $13 \times 71.354\ ms = 0.927\ s$, respectively. Hence, grant-free communication can effectively reduce access latency.

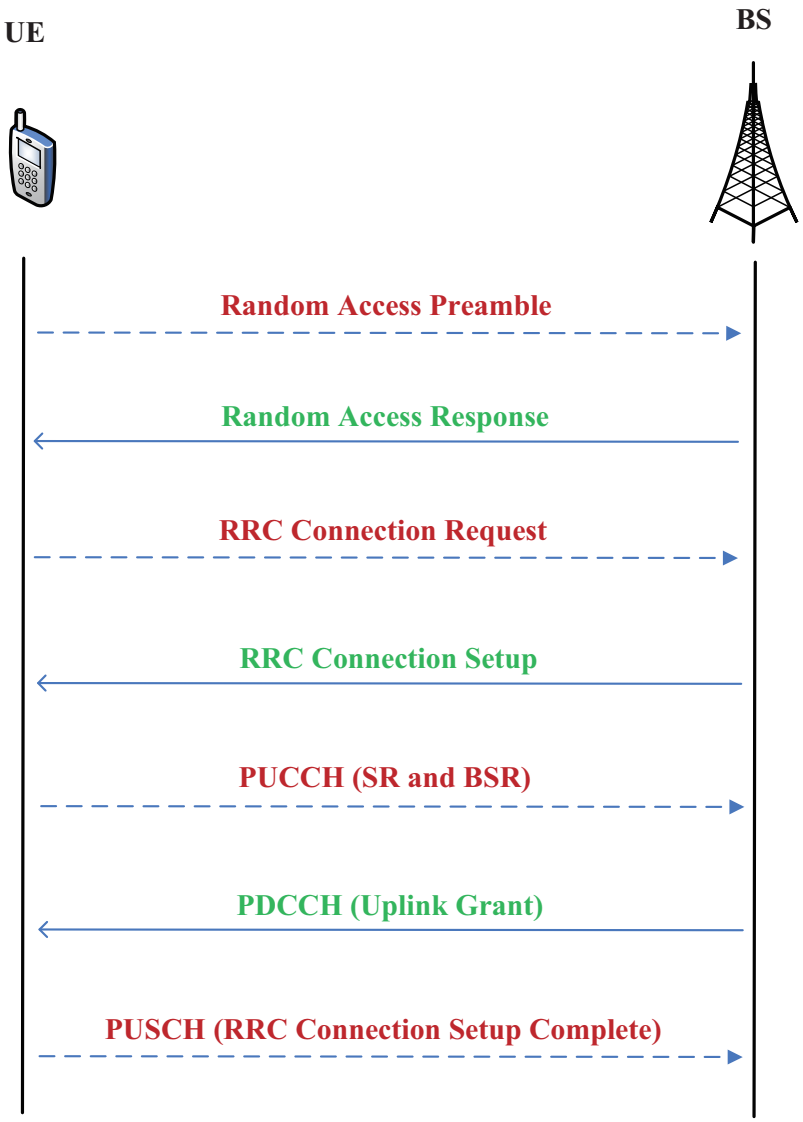

Figure 1.2: RA illustrated in 3GPP.

Table 1.1: Uplink latency.

Steps	**Grant-based (TTI)**	**Grant-free (TTI)**
Minimum waiting time for SR (for grant-based) or transmission opportunity (for grant-free)	1	0
UE sends SR on PUCCH	1	N/A
BS decodes SR and generates the scheduling grant	3	N/A
Transmission of scheduling grant	1	N/A
UE processing delay	3	N/A
Transmission of uplink data	1	1
Data decoding in eNB and ACK/NACK generation	3	3
Total	13	4

1 TTI = 1 ms in LTE or 1 OFDMA symbol in 5G NR. 1 OFDMA symbol = 71.354 ms with 15 kHz subcarrier spacing and normal CP; N/A: Not available

1.1.2 Non-Orthogonal Multiple Access

According to a recent Cisco annual internet report, the number of network devices will reach 29.3 billion by 2023, of which the number of MTC devices (MTCD) is expected to be 14.7 billion; accounting for about 50% of the global connections . Similarly, Ericsson predicts that there will be around 31.4 billion devices to be connected to communication networks by 2023, of which more than 60% will be MTCDs and other IoT connections [22]. Considering this huge number of MTCDs, dramatic changes are needed to the current protocols. This is exacerbated by the fact that the allocation of wireless resources in existing multiple access (MA) techniques, *i.e.*, orthogonal MA (OMA), is inherently non-overlapping and can not allocate wireless resources to massive users. Considering a large number of devices and the limited available radio resources, the OMA-based resource allocation becomes a performance bottleneck [23, 24, 25].

OMA is a realistic choice for achieving good performance in terms of system-level throughput. However, due to the upcoming challenge of massive connectivity, 5G networks need to further improve system performance [26]. In this context, non-orthogonal MA (NOMA) is considered as a promising technology to provide massive connectivity. Based on the non-orthogonal principle, multiple users can transmit their data simultaneously on the same radio resource by using a sequence of user-specific MA signatures, which can be used by the receiver for efficient data recovery. In the last few years, NOMA has attracted a lot of attention from researchers trying to meet the requirements of 5G [27, 28, 29]. Hence, there are many studies in this area. Research trends on NOMA include a variety of topics such as various performance analysis methods, fairness analysis, and user pairing. Many researchers are trying to improve wireless technologies' performance by using NOMA, such as collaborative communications, multiple-input multiple-output (MIMO), optical communications, and relay networks [30, 31, 32].

There are different NOMA solutions, which can be divided into two main approaches, consisting of power-domain NOMA and code-domain NOMA [33]. The power-domain NOMA implements multiplexing in the power domain, while code-domain NOMA implements multiplexing in the code domain. As with basic CDMA systems, code-domain NOMA shares the entire available resource (time/frequency), and user-specific spreading sequences are employed in code-domain NOMA. The spreading sequences are either sparse or non-orthogonal sequences. In fact, the code-domain NOMA can be further divided into several different categories, such as Low-Density Spread Spectrum based OFDM (LDS-

OFDM) [34], Low-Density Spread Spectrum CDMA (LDS-CDMA) [35], and Sparse Code Multiple Access (SCMA) [36]. LDS-OFDM can be thought of as the combination of OFDM and LDS-CDMA, in which information symbols are first expanded over a low-density expansion sequence. Then the corresponding chips are transmitted over a set of subcarriers. SCMA is a recent code-domain NOMA technology based on LDS-CDMA. In contrast to LDS-CDMA, bits in SCMA can be directly mapped to different sparse codewords. SCMA provides a low-complexity reception technique and offers improved performance compared to LDS-CDMA. In fact, other access techniques are closely related to NOMA, including spatial division multiple access (SDMA) [37] and pattern division multiple access (PDMA) [38]. PDMA can be implemented in a variety of domains. PDMA first maximizes diversity and minimizes overlap between multiple users to design non-orthogonal patterns on the transmitter side. Multiplexing is then performed in the spatial domain, the code domain, or a combination of them. SDMA is inspired by the basic CDMA system. Instead of using user-specific spreading sequences, SDMA utilizes user-specific channel impulse response to distinguish between different users.

In a nutshell, to address the limited radio resources and the massive connectivity challenges, NOMA has emerged as a promising technology that allows multiple users to transmit their data simultaneously on the same channel resource. This is achieved by using user-specific signature sequences on the transmission device to implement user activity detection (UAD), channel estimation and signal detection at a receiver.

1.1.3 Compressive Sensing Techniques

To achieve massive connectivity and address the limitations of existing MA solutions, the design of new wireless technologies is inevitable. NOMA has emerged as a potential MA technology for 5G and beyond [39]. The idea behind NOMA is to superimpose multiuser data on the same physical resources by employing a user-specific signature pattern that facilitates burst user detection at the BS. Since data generated by different users in IoT is sporadic, UAD and signal detection are essential. The BS needs to identify K users who transmit data in a specific time slot from T total users. For this reason, the use of compressive Sensing (CS) algorithms for receiver design has been attracted a lot of attention for grant-free communication [40].

Currently, large amounts of information needs to be processed or transmitted, which introduces new challenges, such as large memory densities, high power processing and

increased energy consumption. In some applications, such as radar, imaging, data acquisition, and speech recognition, the signals involved can be considered compressive or sparse in some domains [41]. Since CS can recover sparse signals with fewer measurements than conventional methods, it can be a suitable candidate to deal with these challenges [42].

The existing recovery algorithms can be divided into three parts, including greedy algorithms, convex optimization and non-convex optimization [43]. The core idea of greedy algorithms is to iteratively recover the signal, making locally optimal choices at each iteration to find the optimal global solution at the end of the algorithm. Convex optimization algorithms can recover the signal by advanced techniques such as the interior point method, the iterative thresholding method, or the projected gradient method. On the other hand, non-convex optimization methods is performed by considering prior knowledge of the signal distribution. Thanks to the posterior probability density function, these solutions provide the complete statistics of the estimation.

1.2 Research Motivation

NOMA has emerged as a promising technology to address the challenge of limited radio resources for future wireless communications. The core concept behind NOMA is to manage inter-user interference by superimposing multiuser signals on the same radio resource using dedicated MA signatures, where MA signatures are typically designed in power, code, or interleaved domains. Given this guideline, various techniques have been studied to maximize sum rate under power budget, minimize power under signal-to-interference-plus-noise ratio (SINR) constraint and minimize SINR outage probability, power control, and user grouping through cooperative NOMA. Furthermore, NOMA has been applied to several communication systems, including but not limited to MIMO NOMA, intelligent reflective surface-based NOMA, and wireless power grant-based NOMA. However, the studies mentioned above rely heavily on traditional request-based authorization for user access. The user and the BS need to exchange a large number of signaling messages for user access, thus imposing excessive training overhead and delay on the MTC. To cope with it, grant-free NOMA communication with low overhead and latency is studies. Specifically, each user can transmit data without a strict access authorization process, which is considered as the main performance bottleneck of MTC.

In fact, in the MTC scenario, active users are only a small fraction of the total number of potential users, usually less than 10%, where signals generated by the active user tend

to be sparse in the time domain and have strong sparsity for uplink transmission. For this reason, CS has been a critical component of grant-free NOMA systems as an effective signal processing technique with great potential for sparse signal detection for grant-free communication.

As previously mentioned, NOMA based burst user detection has become a major focus for grant-free communication to facilitate transmission from so many IoT users with very low latency. For this reason, there is a motivation to investigate a deeper relationship between CS and grant-free NOMA systems to provide low-latency access solutions in the context of large-scale connectivity.

1.3 Research Objectives

In light of the above motivation, the aim of this thesis is to provide CS based grant-free communication in the presence of massive potentially active users without moving or in low-speed moving scenarios. As shown in Figure 1.3, the researches of CS based grant-free communication focus on UAD, channel estimation and signal detection in this thesis. To this end, the following research questions are considered.

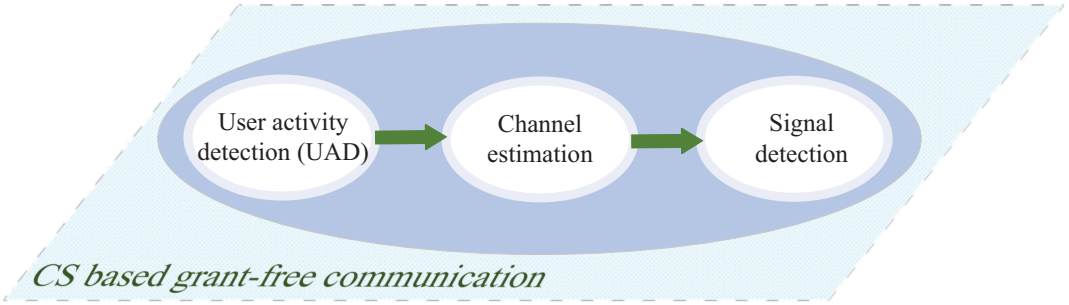

Figure 1.3: The main focus of grant-free communication.

- Is it possible to improve the performance of conventional CS based access solutions by making full use of channel characteristics?
- Periodic communication is one of the salient features of IoT. How the periodic communication can affect the access transmission?

- A practical communications consists of massive users that are actively connected to BS, and a small portion of to-be-connected users. How to ensure the reliability of low-latency access in the presence of massive connected users?
- How to ensure the secure uplink access communication alongside with the requirement of low-latency and massive users?

With these research questions, the objectives of this thesis can be summarized as follows,

- **Grant-free access for sporadic communication**. The relationship between UAD and channel estimation with relatively time-invariant multipath delay is investigated for sporadic communication. In particular, low SNR scenarios is considered since UAD and channel estimation are more challenging in multipath channel compared to single-path channel, especially for low SNR.
- **Grant-free access for periodic communication**. Joint CS based period estimation, UAD, channel estimation and signal detection for low-latency access in the context of periodic communication is investigated.
- **Grant-free access in the presence of massive connected users**. UAD and signal detection for low-latency access in the presence of massive connected users and multiple antennas at the BS is investigated.
- **Channel estimation for secure grant-free communication**. Transmission pilots is investigated in time-division duplex systems to prevent eavesdroppers from acquiring users' channel information.

1.4 Thesis Outline

This thesis summarizes the outcomes of the Ph.D. Project. The main content of the thesis is based on the collection of the papers published during the Ph.D. study. Figure 1.4 illustrates the structure of the thesis, providing a guideline on how the context is related to the reference publications. Note that, for clarity and in the following chapters, the term user is used to refer to different terminals, such as device, UE, MTCD, *etc.*

In **Chapter** 1, the background of the research topic is introduced, alongside with the discussion of motivation and objectives of this thesis.

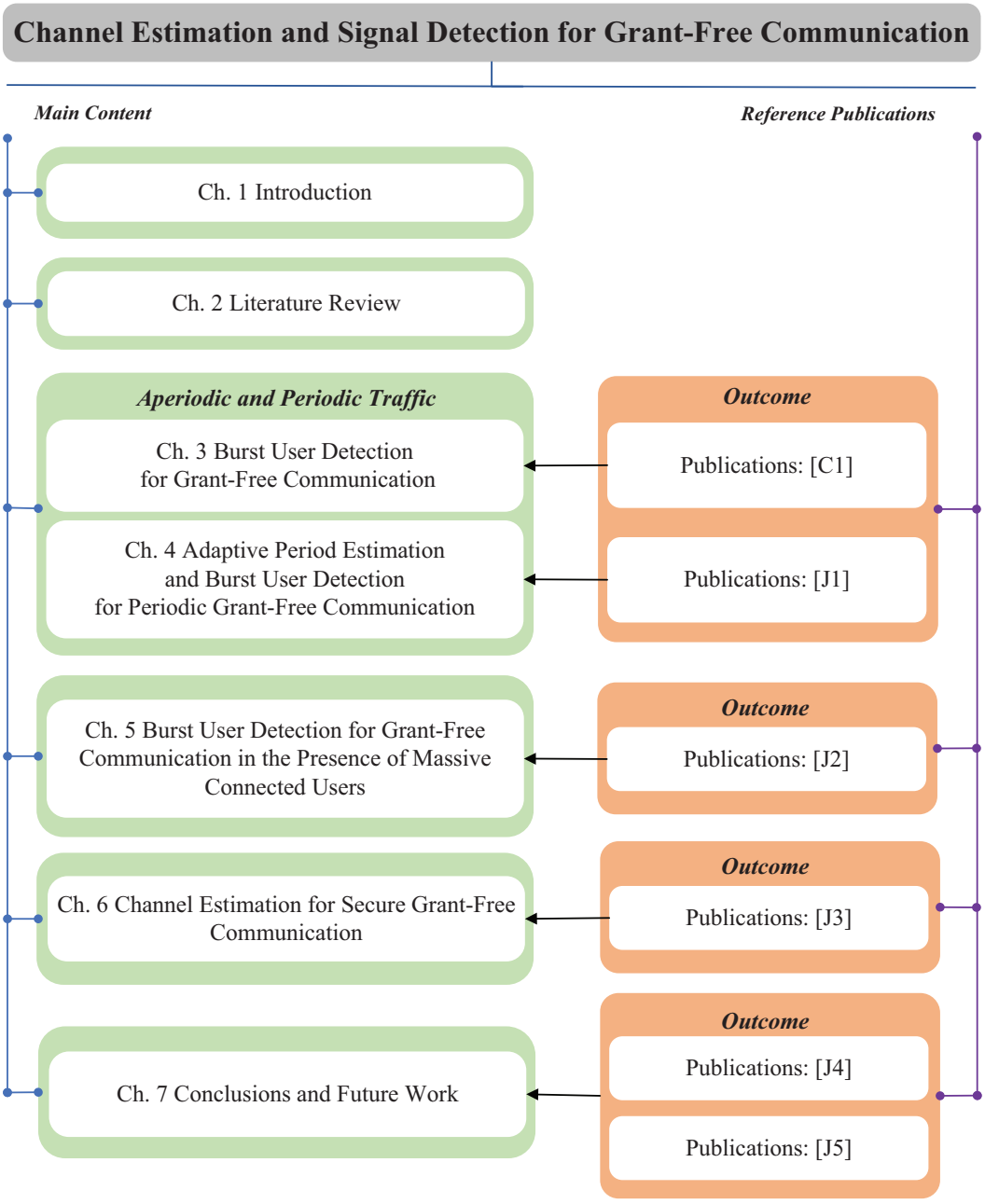

Figure 1.4: Thesis structure.

Chapter 2 provides a general review of user detection for grant-free communication.

In **Chapter** 3, CS aided UAD is investigated to enable grant-free access in NOMA systems, by utilizing the property of quasi-time-invariant channel tap delays as the prior information. The quasi-time-invariant channel tap delays is in line with the scenarios that channel tap delays vary slowly in time domain, which is nearly invariant compared to channel tap gain. Two UAD algorithms are proposed and achieve high UAD accuracy at low SNR.

Chapter 4 introduces UAD schemes in periodic communication, where data is periodically collected from a large number of users and transmitted to an access point (AP). This is the first work to exploit the non-continuous temporal correlation of the received signal for joint UAD, channel estimation and signal detection for grant-free communication. Two schemes are proposed toward this purpose, namely periodic block orthogonal matching pursuit (PBOMP) and periodic block sparse Bayesian learning (PBSBL), which outperform the existing schemes in terms of the success rate of UAD, bit error rate (BER) and accuracy in period estimation and channel estimation. The Cramér-Rao lower bounds (CRLBs) of channel estimation by PBOMP and PBSBL are derived.

Chapter 5 considers a more practical scenario of grant-free communication than **Chapter** 3 and **Chapter** 4, where massive users are actively connected to the BS, while a small portion of to-be-connected users wake up in a burst. The number of effective interfering connected users is reduced to only one, by applying a preconditioning matrix to the received signals from multiple antennas at the BS. Subsequently, an iterative UAD and signal detection scheme is applied, where the *priori* information about the remaining interfering user is exploited and the symbols of the burst users as well as the signature connected user are simultaneously detected by iterative exchanging of the soft information of user activity and symbols.

Chapter 6 focuses on the issue of the physical layer for grant-free communication. The state of the art CS algorithms rely on a very limited category of measurement matrix, *i.e.*, pilot matrix, which may be analyzed by an eavesdropper (Eve) to infer the user's channel information. To this end, the channel reciprocity in time-division duplex systems is utilized to design environment-aware (EA) pilots derived from transmission channels to prevent eavesdroppers from acquiring users' channel information.

Chapter 7 draws the main conclusions of the thesis and discusses the potential future work.

1.5 Publication List

1.5.1 Journal Publications

[J1] **Y. Wang**, X. Zhu, E. G. Lim, Z. Wei and Y. Jiang, "Grant-Free Communications with Adaptive Period for IIoT: Sparsity and Correlation based Joint Channel Estimation and Signal Detection," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4624-4638, 2022.

[J2] **Y. Wang**, X. Zhu, E. G. Lim, Z. Wei, Y. Jiang and L. Gan, "Detection of Burst Users and Symbols for Grant-Free Communication in the Presence of Massive Connected Users," accepted by *IEEE Transactions on Vehicular Technology*.

[J3] **Y. Wang**, E. G. Lim, Y. Zhang, B. Song, R. Pei and X. Zhu, "Compressive Sensing based Secure Uplink Grant-Free Systems," *Frontiers in Signal Processing*, vol. 2, no. 837870, pp. 1-6, 2022.

[J4] **Y. Wang**, E. G. Lim, X. Xue, G. Zhu, R. Pei and Z. Wei, "Interference Utilization Precoding in Multi-Cluster IoT Networks," *Frontiers in Signal Processing*, vol. 1, no. 761559, pp. 1-5, 2021.

[J5] H. Zeng, J. Wang, Z. Wei, X. Zhu, Y. Jiang, **Y. Wang** and C. Masouros, "Next Generation Industrial IoT Networks: the Era of Non-Orthogonal Transmission," accepted by *IEEE Vehicular Technology Magazine*.

1.5.2 Conference Publications

[C1] **Y. Wang**, X. Zhu, E. G. Lim, Z. Wei, Y. Liu and Y. Jiang, "Compressive Sensing based User Activity Detection and Channel Estimation in Uplink NOMA Systems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, South Korea, May 2020, pp. 1-6.

[C2] J. Cao, X. Zhu, Y. Jiang, Z. Wei, S. Sun and **Y. Wang**, "Peak AoI vs. Delay: Closed-Form Analysis in the Finite Block Length Regime," in *13th International Conference on Wireless Communications and Signal Processing (WCSP)*, Changshang, China, Oct. 2021, pp. 1-6.

[C3] Q. Xiong, X. Zhu, Y. Jiang, J. Cao and **Y. Wang**, "Status Prediction for Age of Information Oriented Short-Packet Transmission in Industrial IoT," accepted by *2022 IEEE Wireless Communications and Networking Conference (WCNC)*.

[C4] B. Li, X. Zhu, Y. Jiang, H. Zeng and **Y. Wang**, "Cooperative Time Synchronization and Parameter Estimation via Broadcasting for Cell-Free Massive MIMO Networks," accepted by *2022 IEEE Wireless Communications and Networking Conference (WCNC)*.

[C5] Y. Zhang, X. Zhu, Y. Jiang, Y. Liu and **Y. Wang**, "Hierarchical BEM based Estimation of Doubly Selective Channels for OFDM systems," accepted by *2022 IEEE 95th Vehicular Technology Conference (VTC Spring)*.

[C6] Z. Guo, X. Zhu, Z. Wei, Y. Jiang and **Y. Wang**, "Collision-Aware Random Access Control with Preamble Reuse for Industrial IoT," accepted by *2022 IEEE 95th Vehicular Technology Conference (VTC Spring)*.

Chapter 2

CS and Grant-Free Communication

2.1 Fundamentals of CS

Figure 2.1 provides a framework of CS in a noiseless scenario, a signal $\mathbf{h} \in \mathbb{C}^{N\times 1}$ is said to be K-sparse if it has only K non-zero elements, where $K \ll N$. In this context, $\mathbf{h}$ can be projected into a lower dimensional space, referred to as a compressible signal $\mathbf{y} \in \mathbb{C}^{M\times 1}$. For a noisy scenario, the system is defined by

$$\mathbf{y} = \mathbf{\Theta h} + \mathbf{z} \tag{2.1}$$

where the matrix $\mathbf{\Theta}$ is referred to as measurement matrix and $\mathbf{z}$ indicates the noise vector. By employing CS theory, $\mathbf{h}$ can be recovered by $\mathbf{y}$ if $K < M$ and $M < N$.

The key challenge of CS is to recovery the unknown signal $\mathbf{h}$. This could be achieved by identifying the locations of the nonzero signal components in $\mathbf{h}$. In detail, it is to find the subspace generated by no more than K columns of the matrix $\mathbf{\Theta}$, related to the received signal $\mathbf{y}$. After finding these locations, the nonzero coefficients can be obtained by employing a pseudo-inversion process. In this context, CS theory addresses two main problems,

- Design of the measurement matrix $\mathbf{\Theta}$.
- Development of a sparse recovery algorithm for $\mathbf{h}$ estimation, given only $\mathbf{y}$ and $\mathbf{\Theta}$.

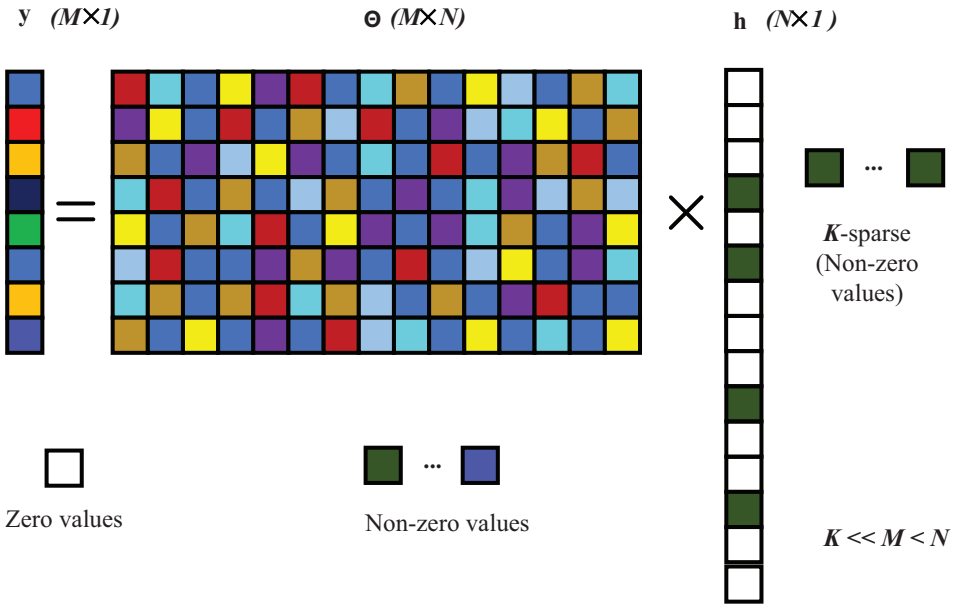

Figure 2.1: Representation of measurements used in compressive sensing.

The goal of designing a measurement matrix $\mathbf{\Theta}$ is to ensure that the main information of any K sparse signal is in this matrix. The measurement matrix is very important in the sparse signal recovery process. According to [43, 44], if the restricted isometric property (RIP) defined in (2.2) is satisfied, it is possible to obtain an accurate estimate of the sparse signal $\mathbf{h}$ using some recovery algorithm. $\delta \in (0, 1)$ is the RIC (restricted isometric constant) value corresponding to the smallest number that reaches (2.2).

$$(1 - \delta) \|\mathbf{h}\|_2^2 \leq \|\mathbf{\Theta h}\|_2^2 \leq (1 + \delta) \|\mathbf{h}\|_2^2 \tag{2.2}$$

Currently, the measurement matrices of CS for grant-free communication are mainly generated by identical and independent distributions (i.i.d.) such as Bernoulli, Gaussian, and random Fourier ensembles. For the second problem, CS based burst user detection have been extensively investigated for grant-free communication [45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55]. These schemes can be categorized into three structures: (1) sporadic sparsity structure, (2) burst-sparsity structure, and (3) frame-wise joint sparsity structure, as detailed in the following section.

2.2 CS based Burst User Detection

For grant-free NOMA communication, each active device transmits its information directly to the BS without waiting for any permission, as shown in Figure 2.2. In contrast to the grant-based random access scheme where the pilot is randomly selected at each time slot, each device in grant-free communication is preassigned with a unique sequence for all time slots, where sequence also serves as the user's identification [18, 56, 57, 45]. The BS first detects the active device at each time slot by detecting which pilots are used. Then, the BS estimates their channels based on the received metadata and decodes the data using the estimated channels.

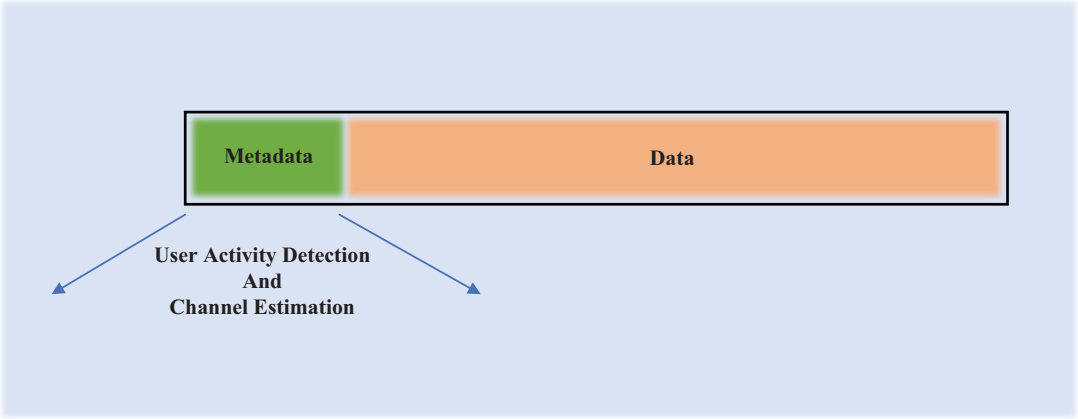

Figure 2.2: A grant-free transmission strategy. Metadata contains preamble for user activity detection (UAD) and channel estimation, and data is directly transmitted after metadata without waiting for the grant from the BS.

In light of the sparsity of the active users in IoT, CS emerges as an effective signal processing technique for grant-free transmission to perform UAD, channel estimation and signal detection [22, 58, 59, 60, 61]. CS exploits the sparsity of a signal to recover it from far fewer samples than required by Nyquist criteria [43]. In this section, a general review of CS based burst user detection for grant-free communication is illustrated in Subsection 2.2.1. Then the assumption of the existing CS based grant-free communication are discussed in Subsection 2.2.2, 2.2.3, 2.2.4 and 2.2.5, respectively.

2.2.1 Three Structures for CS based Grant-Free Communication

In the early stage, most studies of CS based grant-free communication focused on the standard CS model, *i.e.*, the sporadic transmission without the consideration of temporal correlation [47]. This sporadic structure shown in Figure 2.3, corresponds to the scenario where users transmit their data randomly in each time slot.

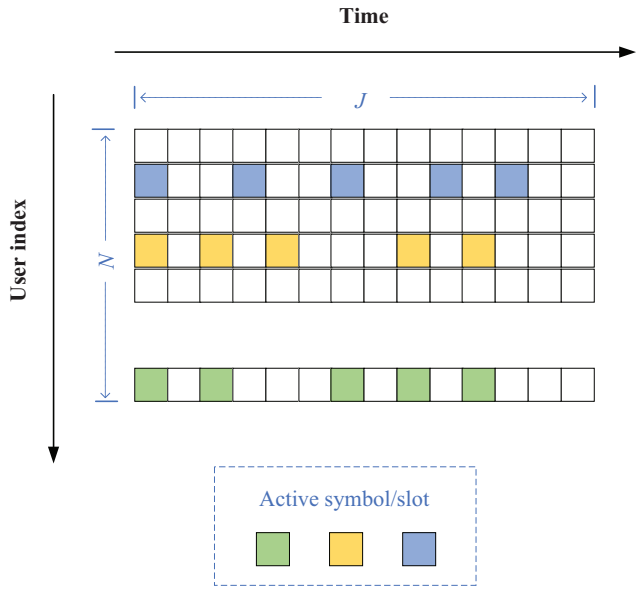

Figure 2.3: Representation of sporadic sparsity structure.

Based on the sporadic sparsity structure of CS, the authors of [62] considered an burst user detection when the sparsity of the signal varies with time. Then a switching scheme was proposed, where the classical orthogonal matching pursuit (OMP) and linear minimum mean squared error (LMMSE) detector are self-adaptively switched to perform burst user detection. In general, the sparse signal is independently reconstructed in the context of sporadic sparsity structure.

In fact, the data generated from different users generally is transmitted in several continuous time slots. Hence, the user sets can be detected by exploiting the correlation between the continuous data in adjacent time slots. This temporal correlation makes more robust CS based techniques possible, which is detailed later in this section. By exploiting the correlation between the continuous data in adjacent time slots, there are two sparsity

models, *i.e.*, burst-sparsity structure and frame-wise joint sparsity structure [48].

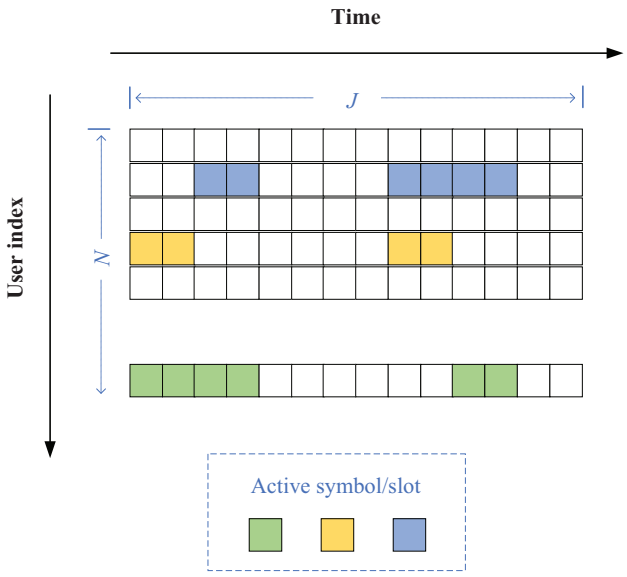

Figure 2.4: Representation of burst-sparsity structure.

As shown in Figure 2.4, burst-sparsity structure is in line with the scenario where some users generally transmit their information in adjacent transmission symbols with a high probability. Given this guideline, a dynamic CS based (DCS) UAD was designed to realize data detection aided by the temporal correlation of the active user sets [47]. The idea behind DCS-based UAD is that the active user sets between adjacent time slots are considered as the prior information, which can be used to estimate the active user set over entire slots [49]. The idea of temporal correlation between adjacent time slots was also exploited in [50] to enhance the performance of UAD detection in the multiple-input multiple-output (MIMO) BS. Different from [47] which requires the knowledge of the number of active users, the work in [48] utilizes the quality information to adaptively obtain the user sparsity level. Furthermore, two prior-information aided adaptive subspace pursuit (PIA-ASP) algorithms was proposed, and the simulation results showed the PIA-ASP algorithms outperform the DCS algorithms in terms of UAD performance.

Figure 2.5 presents the frame-wise joint sparsity structure, where activity and inactivity of users keep unchanged over an entire data frame. Based on the frame-wise joint sparsity structure, the authors of [51] demonstrated a low complexity block sparse Bayesian

learning (SBL) scheme combined with message passing to perform UAD and channel estimation. It is showed that the complexity of the proposed message passing based block SBL (MP-BSBL) algorithm is independent to number of active user, and hence, the MP-BSBL can be significantly lower than that of the greedy algorithms, *e.g.,* block orthogonal matching pursuit. Similarly, the block subspace pursuit algorithm [52], group orthogonal matching pursuit algorithm [53], enhanced subspace pursuit algorithm [54], block compressive sampling matching pursuit (BCoSaMP) algorithm [63], independent component analysis sparse Bayesian learning (ICASBL) algorithm [64], *etc.*, were proposed to perform UAD, channel estimation or signal detection. Particularly, an iterative order recursive least square (IORLS) algorithm was proposed to exploit the frame-wise joint sparsity structure, which can increase accuracy of UAD by gathering the sparsity support information with manageable complexity. Generally speaking, the above mentioned frame-wise joint structure algorithms were developed from the theory of block CS and achieved an excellent performance of UAD.

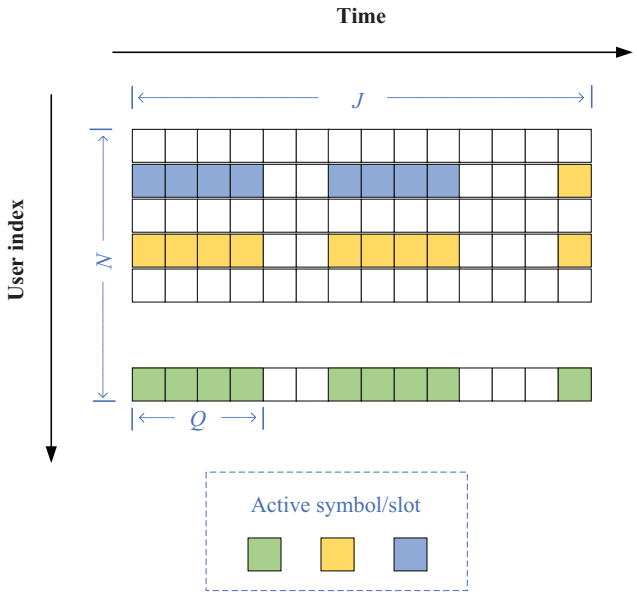

Figure 2.5: Representation of frame-wise structure.

2.2.2 Stop Condition and Channel

An enormous amount of CS based research effort goes into sporadic sparsity structure, burst-sparsity structure and frame-wise joint sparsity structure. For example, an DCS-based algorithm is proposed to implement burst user detection by combining the OMP algorithm with the temporal correlation of the active user sets [47]. The OMP algorithm recover the signal in an iterative way, making a local optimal selection at each iteration hoping to find the global optimum solution at the end of the algorithm. In detail, let $\hat{\mathbf{h}}_0 = 0$, each iteration i of the OMP algorithm consists in finding the column $\mathbf{a}_{k,i} \in \boldsymbol{\Theta}$ in (2.1), which is the best aligned with the residual vector $\mathbf{b}_i(\mathbf{b}_0 = \mathbf{y})$. Building upon this structure, the index k_i can be obtained by

$$k_i = \arg\max_{l} |\mathbf{a}_l^{\mathrm{H}} \mathbf{b}_{i-1}|, \quad l = 1, 2, \cdots, N. \tag{2.3}$$

The index set Λ_i stores the indices of the best aligned with columns after i iterations. The index k_i is merged into the index set Λ_i, *i.e.,*

$$\Lambda_i = \Lambda_{(i-1)} \cup k_i. \tag{2.4}$$

Subsequently, a new residual vector is computed as

$$\mathbf{b}_i = \mathbf{y} - \boldsymbol{\Theta}(\Lambda_i)\hat{\mathbf{h}}_i, \tag{2.5}$$

where

$$\hat{\mathbf{h}}_i = \boldsymbol{\Theta}^{\dagger}(\Lambda_i)\mathbf{y}. \tag{2.6}$$

The OMP algorithm is mainly implemented by (2.3), (2.4), (2.5) and (2.6). The algorithm is terminated if $i = K$. However, the number of active users, *i.e.,* the sparsity K can not be obtained for grant-free communication. Hence, the requirement of knowledge of sparsity of active users hinders CS application into practice. To this end, the authors of [49] obtained the number of active users by estimating the noise level. Nevertheless, the works in [47] and [49] assumed single path channel. In fact, compared to single-path channel, superimposed signals from multiple users may be distorted by the multipath effect, especially for low SNR scenario. Hence, CS based grant-free communication should be investigated in a practical multipath channel.

2.2.3 Correlation between the Continuous Data

As presented in Subsection 2.2.1, there has been a lively interest in the CS based grant-free communication based on the burst-sparsity structure and the frame-wise joint sparsity structure by exploiting the correlation between the continuous data in adjacent time slots. For example, the authors of [51] demonstrated a low complexity block SBL scheme by combining SBL and the temporal correlation among the continuous data. To illustrate this, we first give the concept of SBL. Let σ^2 be the noise variance, the SBL assumes the Gaussian likelihood model,

$$p\left(\mathbf{y} \mid \mathbf{h}; \sigma^2\right) = \left(2\pi\sigma^2\right)^{-M/2} \exp\left(\frac{-1}{2\sigma^2}||\mathbf{y} - \mathbf{\Theta h}||^2\right) \tag{2.7}$$

In light of the Bayesian probability theory, a class of prior probability distributions $p(\theta)$ is conjugate to a class of likelihood functions $p(\theta|x)$ if the resulting posterior distributions $p(x|\theta)$ are in the same family as $p(\theta)$. Since the Laplace prior is not conjugate to the Gaussian likelihood, the relevance vector machine is used. Based on (2.7), the solution of (2.1) can be achieved by a type-II maximum likelihood procedure and Expectation-Maximization algorithm, as detailed in Chapter 4. The SBL does not consider the correlation between the continuous data. While block SBL exploited the correlation by converting multi-dimension vector to single vector based on Kronecker product. For example, it is assumed that $\mathbf{Y} = [\mathbf{y}_1^{\mathrm{T}}, \mathbf{y}_2^{\mathrm{T}}, \cdots, \mathbf{y}_J^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{H} = [\mathbf{h}_1^{\mathrm{T}}, \mathbf{h}_2^{\mathrm{T}}, \cdots, \mathbf{h}_J^{\mathrm{T}}]^{\mathrm{T}}$ and J indicates the number of received symbols. Accordingly, we can arrive at

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{h}}\widetilde{\mathbf{\Theta}} + \widetilde{\mathbf{z}}, \tag{2.8}$$

where $\widetilde{\mathbf{y}} = \mathrm{vec}(\mathbf{Y})$, $\widetilde{\mathbf{\Theta}} = \mathbf{\Theta}^{\mathrm{T}} \otimes \mathbf{I}$, $\widetilde{\mathbf{h}} = \mathrm{vec}(\mathbf{H})$, $\widetilde{\mathbf{z}} = \mathrm{vec}(\mathbf{Z})$, and $\mathbf{I} \in \mathbb{C}^{J \times J}$ is an identity matrix. In this context, (2.7) can be employed. Accordingly, the block SBL has improved UAD performance resort to the correlation between the continuous data.

Technically speaking, the requirement of continuous data in adjacent time slots may not be satisfied. For example, periodic communication is one of the salient features of IoT, where data is periodically collected from a large number of users and transmitted to an AP [9, 65]. Furthermore, the transmission period of various users may adaptively change due to the specific requirements of energy-saving or complex control [66, 67]. In general, the existing CS based schemes do not involve periodic communication, in which the transmission period of various users may adaptively change. Hence, intrinsic information

(*e.g.*, non-continuous data temporal correlation among the periodic communication) is not considered in the existing CS based schemes [47, 49, 50].

2.2.4 Massive-User System

The existing assumption of massive-user system for grant-free communication is that there are only burst users and the system is defined by

$$\mathbf{y} = \mathbf{H}_{\mathbb{S}_\mathrm{A}} \mathbf{\Theta}_{\mathbb{S}_\mathrm{A}} \mathbf{x}_{\mathbb{S}_\mathrm{A}} + \mathbf{z}, \tag{2.9}$$

where the under-scripts $\mathbb{S}_\mathrm{A}$ denote the set of the burst users and $\mathbf{x}$ denotes the transmission data. For example, in [47], CS was used to detect the users who are active in scattered time slots without users that are actively connected to the BS. Considering a more practical consecutive time slots transmission, a maximum of *posterior* probability based approach was proposed for UAD by exploiting the temporal correlation in the adjacent time slots [68]. In [55] and [69], CS based UAD schemes were proposed for the cases of aperiodic and periodic traffic, respectively, by employing the temporal correlation of the received signals. However, the aforementioned work [47, 55, 68, 69] assumed that BS is equipped with a single antenna.

To be in line with a practical multi-antenna case, an OMP algorithm was developed for the low-complexity channel estimation in massive MIMO systems [70], and an efficient matching pursuit (EMP) algorithm was proposed for UAD [71]. Nevertheless, the inherent advantage of multiple antennas on CS designing has not been fully exploited. More importantly, all the aforementioned work has not considered the impact of massive connected users on UAD for grant-free communication. Hence, considering of massive BS antennas and massive connected users, (2.9) is rewritten as

$$\mathbf{y} = \mathbf{H}_{\mathbb{S}_\mathrm{C}} \mathbf{\Theta}_{\mathbb{S}_\mathrm{C}} \mathbf{x}_{\mathbb{S}_\mathrm{C}} + \mathbf{H}_{\mathbb{S}_\mathrm{A}} \mathbf{\Theta}_{\mathbb{S}_\mathrm{A}} \mathbf{x}_{\mathbb{S}_\mathrm{A}} + \mathbf{z}, \tag{2.10}$$

where the under-scripts $\mathbb{S}_\mathrm{C}$ denotes the set of the connected users. Clearly, CS based grant-free communication in (2.7) is more challenge than that of (2.9) due to the interference of massive connected users. Hence, CS based grant-free communication should be investigated in the presence of massive connected users and multiple antennas at the BS.

2.2.5 Pilot Matrix

The existing pilot matrix, *i.e.,* the measurement matrix $\mathbf{\Theta}$ in (2.1), of CS based grant-free communication is mainly marked by Bernoulli distribution [55] and Gaussian distribution [72]. As described in Section 2.1, one of the challenges for CS is to design the measurement matrix. In [73], the geometrical structure of the BS's antenna array is utilized to design the non-orthogonal pilot for channel estimation, where the pilot matrix is equivalent to the discrete Fourier transform (DFT) matrix in the virtual angular domain. In [69], Bernoulli based pilot in the time domain was proposed for channel estimation of active users for grant-free communication. In light of that the majority of users transmit their signals during consecutive slots in practice, the temporal correlation of the active user sets between adjacent slots is employed to design the pilot sequence to enhance the performance of CS based grant-free communication. Given this guideline, the authors of [74] investigated the relationship between the block-sparsity of the uplink access signals and the improved CS based channel estimation, where the pilot matrix keeps unchanged over the entire transmission period.

Currently, the existing CS based channel estimation techniques rely on the specific pilot matrix for grant-free communication. The constant pilot matrix may introduce serious security problems when an eavesdropper has the knowledge of the pilot information. In this context, the physical layer (PHY) security draw lots of attention [5, 6]. Researches in [75] showed that the CS based encryption does not achieve Shannon's definition of perfect secrecy, but a computationally unbounded adversary can easily infer that the correct sparse signal has been recovered. Hence, there are still issues left to be addressed in the security of pilot design for grant-free communication.

2.3 Summary

In light of the increasing of IoT services, there is a need for further new designs, studies and integration of advanced technologies to deal with various challenges, *e.g.,* low complexity schemes, the high accuracy of UAD, secure transmission and *etc.*. In detail and refer to Section 2.3, the main challenges of CS based grant-free communication is summarized as follows.

- UAD and channel estimation are more challenging in multipath channel compared to single-path channel, especially for low SNR. Hence, one of the challenges for grant-

free communication is to design UAD and channel estimation in a practical multipath channel without knowledge of the number of active users, as detailed in Chapter 3.

- How to perform UAD with adaptive period estimation by exploiting non-continuous temporal correlation of the received signals is the key challenge in periodic communication, as detailed in Chapter 4.
- The number of the connected users is usually much larger than the number of to-be-connected burst users, imposing strong interference in UAD and signal detection. Hence, to design the UAD approach in the presence of massive connected users is an indispensable part of CS based grant-free communication, as detailed in Chapter 5.
- The existing CS based channel estimation techniques depend on the specific pilot matrix. Hence, how to design secure CS based channel estimation is one of the challenges of user detection for grant-free communication, as detailed in Chapter 6.

Chapter 3

Burst User Detection for Grant-Free Communication

3.1 Introduction

The strong sparsity of the multipath channel in time domain enables a sparsity feature at uplink transmission. Since CS is able to reconstruct a sparse physical signal with less information [43], it is motivated to apply CS for UAD in NOMA systems, to enable a grant-free communication. As described in Chapter 2, UAD and channel estimation are more challenging in multipath channel compared to single-path channel especially for low SNR. In fact, channel tap delays vary slowly in time domain, which is nearly invariant compared to channel tap gain [76, 77] and is referred to as time-invariant multipath delay. If the characteristic can be effectively exploited as the prior information for UAD, *i.e.,* the total number of active users, can be significantly reduced. Hence, the relationship between UAD and channel estimation with relatively time-invariant multipath delay is investigated in this chapter. The main contributions are summarized as follows.

- To the best of my knowledge, this is the first work to utilize the property of the time-invariant channel tap delays as a prior information to perform CS aided UAD in a NOMA system. The proposed UAD scheme does not require the knowledge of sparsity of active users like the existing work [47, 52, 62, 78], and therefore it is more practical. In particular, it is more suitable for low SNR scenarios than the work in [49], as utilization of the prior information could reduce the influence of noise on CS

aided UAD.

- Two algorithms are proposed for uplink UAD, referred to as gradient based and time-invariant channel tap delays assisted CS (g-TIDCS) and mean value based and TIDCS (m-TIDCS), respectively. The former aims to detect the users based on the fact that the accumulative number of being detected of different active users should be similar during each iteration, while the latter is to detect the users whose accumulative number of being detected is greater than the average number of being detected of all the users. Both algorithms achieve much higher accuracy than the dynamic adaptive CS (DACS) algorithm in [49]. m-TIDCS is more suitable for high active user ratio and very low SNR scenarios, while g-TIDCS is more suitable at a larger-valued SNR (still in the range of low SNR).

- Based on the UAD results obtained, a low-complexity CS based channel estimation approach is proposed, which presents a lower normalized mean square error (NMSE) than the CoSaMP approach in [79] and the OMP approach in [80]. Furthermore, the upper bound and lower bound of complexity ratio between the proposed scheme and OMP[80] are derived, proving a lower complexity of the proposed channel estimator than OMP.

The rest of this section is organized as follows. Section 3.2 illustrates the system model. The detection of set of active user algorithm, together with channel estimation is demonstrated and discussed in section 3.3. Simulation results are given in section 3.4. Finally, conclusions are drawn in section 3.5.

3.2 System Model

A practical uplink NOMA system is considered, where the exact number of active users is unknown at the BS, as shown in Figure 3.1. The users are kept active or inactive in several consistent symbols and each symbol for different users occupies M same subcarriers. Furthermore, there are one BS and N users, all equipped with a single antenna without loss of generality. The number of active users, the maximum path delay and the number of path for each user are denoted as K, L_{M} and L, respectively. h_n^j denotes the time-domain discrete channel vector in the j-th symbol for n-th user with L dominant paths, which is given by $\mathbf{h}_n^j = [0, \cdots, h_n^j(1), 0, \cdots, 0, h_n^j(L), \cdots, 0]^T$. Accordingly, the l-th channel path is

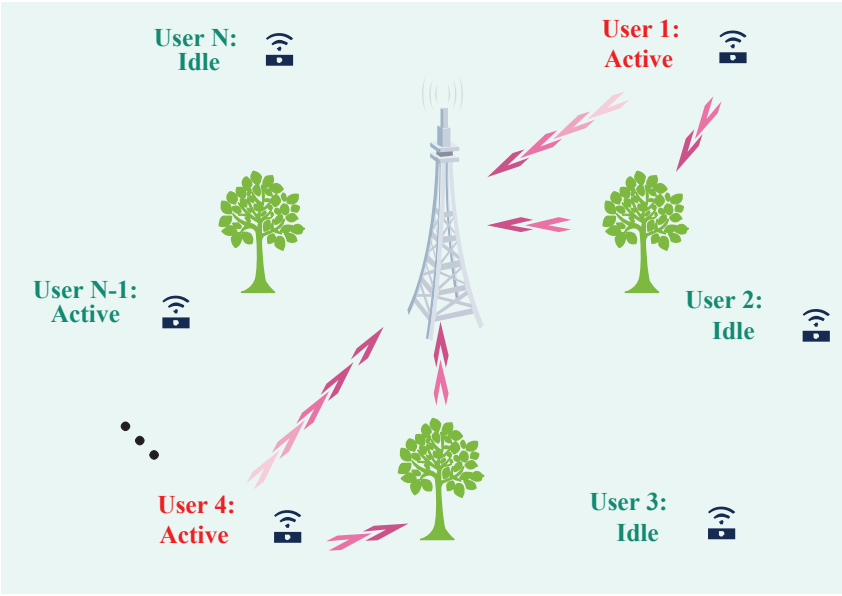

Figure 3.1: Illustration of a typical multipath uplink NOMA scenario.

modeled as

$$h_n^j(l) = \sum_{l=0}^{L-1} \alpha_{n,l}^j \delta(\tau - \tau_{n,l}), \tag{3.1}$$

where $\tau_{n,l}$ and $\alpha_{n,l}^j$ indicate the path delay and the complex path gain of the l-th path, respectively. Particularly, $\alpha_{n,l}^j = 0$ if the n-th user is not active. Furthermore, low-speed moving users are considered and the channels are assumed under the block-fading, meaning that the channel matrix is invariant in a transmission block. Hence, it is assumed that the path delays change slowly and maintain constant for a number of symbols, while the path gains vary from symbol to symbol. Particularly, the initial channel information, including path delay, can be readily estimated from the uplink reference signals. In this context, the path delay can be considered as a prior information to perform CS for grant-free communication.

The received signal for the j-th symbol at the BS can be modeled as

$$\mathbf{y}^j = \sum_{n=1}^{N} \mathbf{\Theta}_n \mathbf{h}_n^j + \mathbf{z}^j, \tag{3.2}$$

where $\mathbf{\Theta}_n = \mathbf{s}_n \mathbf{F}_{L_\mathrm{M}}$ denotes the sensing matrix for the n-th user. $\mathbf{s}_n$ denotes the reference

signal for the n-th user [47]. $\mathbf{F}_{L_{\mathrm{M}}}$ indicates DFT matrix with first L_{M} columns. Furthermore, $\mathbf{z}^j$ denotes the additive white Gaussian noise (AWGN) following complex Gaussian distribution.

For brevity, (3.2) is collected into a compact form as

$$\mathbf{y}^j = \mathbf{\Theta}\mathbf{h}^j + \mathbf{z}^j, \tag{3.3}$$

where $\mathbf{\Theta} = [\mathbf{\Theta}_1, \cdots, \mathbf{\Theta}_N]$ with a size of $M \times NL_M$ and $\mathbf{h}^j = \left[(\mathbf{h}_1^j)^T, \cdots, (\mathbf{h}_N^j)^T\right]^T$ with a size of NL_M. Typically, $M \ll NL_M$, (3.3) is in a standard CS structure where the CS theory can be employed to acquire CSI.

3.3 TIDCS Algorithm

The greedy algorithm is the most popular technique among CS algorithms due to its low complexity. However, it needs to know the sparsity of the measured signal matrix, which is inefficient for grant-free communication. In this section, time-invariant multipath delay is introduced to perform UAD and channel estimation based on OMP [80]. In particular, the proposed approach does not require the knowledge of the level of user activity.

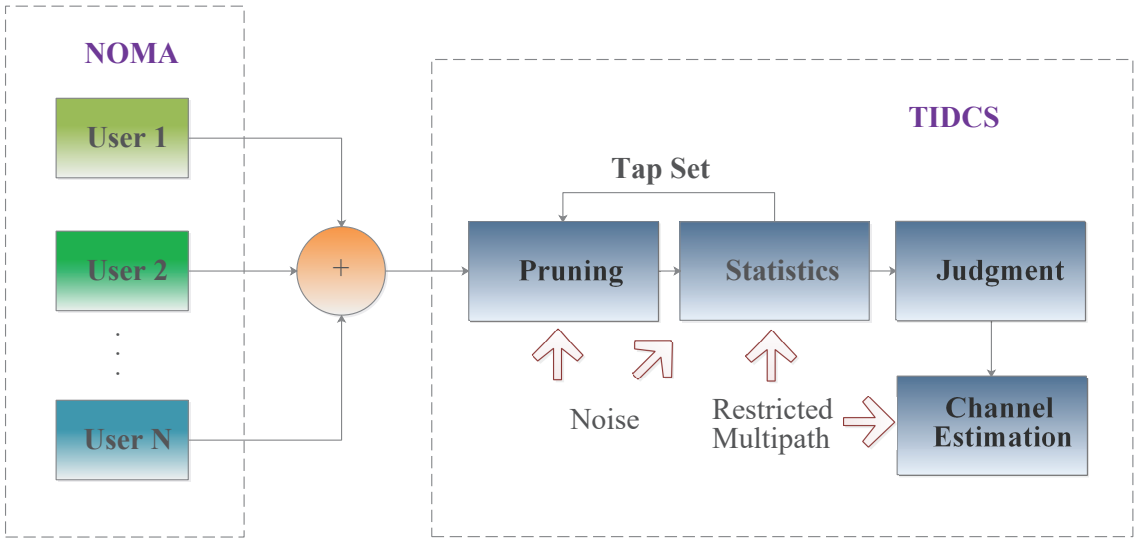

Figure 3.2: Framework for the proposed TIDCS Algorithm.

Figure 3.2 shows the framework for the proposed TIDCS algorithm. The superimposed

received signal from different users is used to perform UAD and channel estimation. The UAD consists of three stages, named as pruning, statistics and judgment, respectively. In particular, noise and time-invariant multipath delay are the prior knowledge of the proposed TIDCS algorithm. Furthermore, channel estimation will be executed after UAD. The detailed of the proposed TIDCS is demonstrated in the following subsection.

3.3.1 User Activity Detection

The traditional greedy algorithm needs to know the exact sparsity level of the received superimposed signal, which is difficult to be acquired for grant-free communication. To cope with it, the residual noise level is introduced to the traditional greedy algorithm to obtain the number of active users. The proposed approach is described as follows and illustrated in **Algorithm 1**.

Pruning

Pruning aims to eliminate erroneously selected channel taps leveraging from previous NOMA symbol. In each NOMA symbol, the selected set of channel tap may contain uncertain value. Hence, currently received data is used to refine the former tap set and correct the recorded set of active user. Furthermore, the power of residual signal is used to determine whether this stage would be performed. The residual signal is given by

$$\mathbf{r}^{j}_{\Omega^{j-1}} = \mathbf{y}^{j} - \Phi_{\Omega^{j-1}} \hat{\mathbf{h}}^{j}_{\Omega^{j-1}}, \tag{3.4}$$

where $\mathbf{r}^{j}_{\Omega^{j-1}}$ is the residual signal in the $j-$th symbol with channel tap set P from the $(j-1)-$th symbol. Note that the power of residual signal equals to the power of current noise if the channel tap set is well recovered, *i.e.,*

$$\left\| \mathbf{r}^{j}_{\Omega^{j-1}} \right\|_2^2 = \left\| \mathbf{z}^{j} \right\|_2^2 . \tag{3.5}$$

Equation (3.5) is employed to assess the former channel tap list Ω^{j-1}. When the power of residual signal is no less than the noise power, redundant tap needs to be pruned. Since the channel tap may be falsely chosen, the impact of false drop of channel tap should be eliminated. Subsequently the index of channel tap ω and corresponding user index n will be collected by step 10 and step 11 as shown in **Algorithm 1**, respectively.

Algorithm 1 TIDCS for UAD

Input:

Received signals: $\mathbf{y}^1, \mathbf{y}^2, \cdots, \mathbf{y}^J$;

Sensing matrix: $\mathbf{\Theta}$.

Output:

User activity set: I_C;

Initialization :

$\Omega = \varnothing; \Upsilon = \varnothing; B = \mathbf{0}; \mathbf{\Psi} = \mathbf{\Theta}$.

for $j = 1$ to J **do**

$\mathbf{\Theta} = \mathbf{\Psi}$;

Pruning:

if $j > 1$ **then**

$\hat{\mathbf{h}}^j_{\Omega^{j-1}} = \mathbf{\Theta}^{\dagger}_{\Omega^{j-1}} \mathbf{y}^j$;

$\mathbf{r}^j_{\Omega^{j-1}} = \mathbf{y}^j - \mathbf{\Theta}_{\Omega^{j-1}} \hat{\mathbf{h}}^j_{\Omega^{j-1}}$;

while $\left\| \mathbf{r}^j_{\Omega^{j-1}} \right\|_2 \leq \left\| \mathbf{z}^j \right\|_2$ **do**

$\omega = \arg\min_{\Omega^{j-1}} \left(\left| \hat{\mathbf{h}}^j_{\Omega^{j-1}} \right| \right)$;

$n = \{\text{User index correspond to } \omega\}$;

$\Omega^{j-1} = \Omega^{j-1} \backslash \{\omega\}$; $B(n) = B(n) - 1$;

$\mathbf{r}^j_{\Omega^{j-1}} = \mathbf{y}^j - \mathbf{\Theta}_{\Omega^{j-1}} \hat{\mathbf{h}}^j_{\Omega^{j-1}}$;

end while

end if

Statistics:

for $i = 1$ to M **do**

$k = \arg\max_k \left| \theta_k^{\mathrm{H}} \mathbf{y}^j \right|$;

$n = \{\text{User index correspond to } k\}$;

if $k > (\Delta_n + \delta)$ or $k < (\Delta_n - \delta)$ **then**

$\mathbf{\Theta}_n = \mathbf{0}$; **Continue**;

end if

$\theta_k = 0$; $\Omega^j = \Omega^j \cup \{k\}$; $\Gamma(n) = \Gamma(n) + 1$;

$\mathbf{h}_{\Omega^j} = \mathbf{\Psi}^{\dagger}_{\Omega^j} \mathbf{y}^j$;

$\mathbf{r}^j_i = \mathbf{y}^j - \mathbf{\Psi}_{\omega^j} \mathbf{h}^j_{\Omega^j}$;

if $||\mathrm{r}^j_i||_2^2 \leq ||\mathrm{z}^j||_2^2$ or $||\mathrm{r}^j_i||_2^2 \leq ||\mathrm{r}^j_{i-1}||_2^2$ **then**

Quit the iteration;

end if

end for

$\Upsilon = \text{find}(\Gamma >= Th)$; $B(\Upsilon) = B(\Upsilon) + 1$;

end for

Judgment:

$v = \max(\text{gradient } (\text{sort}(B)))$ or $v = \text{mean}(B)$;

$I_c = \{\text{Indices for values ne less than } v \text{ in } B\}$;

return I_c.

Statistics

This stage extracts the possible user channel tap and active user index. The misoperation of index along with largest absolute value may occur in low SNR scenario due to noise. Hence, time-invariant multipath delay is employed to confirm step 18. If channel tap delay corresponding to the index k selected by step 18 is beyond the possible range of multipath delay, this index will be discarded. The possible multipath delay has a range from $[\Delta_n - \delta]$ to $[\Delta_n + \delta]$ in current symbol, where the initial set of multipath delay for user n is defined as Δ_n and δ denotes its deviation.

When the candidate tap coincides with set of multipath delay, the candidate tap set is updated by merging the preliminary set and new tap together, *i.e.,*

$$\Omega^j = \Omega^j \cup k. \tag{3.6}$$

Particularly, the column of the sensing matrix corresponding to new tap should be set to $\mathbf{0}$. Furthermore, current set of active user Γ is added for the final judgment. Refer to (3.5), the power of residual signal with perfect support tap set should equal to the power of noise. Accordingly, the part of the statistic is terminated under two conditions. The first condition is based on that the power of residual signal should be no more than the power of noise, while the other one refers to that the power of the current residual signal should be no larger than that of the former one, *i.e.,*

$$||\mathrm{r}_i^j||_2^2 \leq ||\mathrm{z}^j||_2^2 \ or \ ||\mathrm{r}_i^j||_2^2 \leq ||\mathrm{r}_{i-1}^j||_2^2. \tag{3.7}$$

Judgment

For the third stage, judgment is to evaluate the statistical alternative utilizing g-TIDCS or m-TIDCS. These two approaches are presented to verify the final active user. The g-TIDCS is in line with that the accumulative number of being detected for different active users should be similar during each iteration. The collection of B is first sorted, then it can be obtained that

$$\nabla \hat{B} = \frac{\partial \hat{B}}{\partial x}, \tag{3.8}$$

where x denotes the step length and $\hat{B}$ denotes the collection after sorting B. Since the number of being detected for different active users is similar and maximum, the users whose number of times being detected is no less than maximum gradient in B are considered as

the real active users. The m-TIDCS is based on that the number of times being detected for different active users should be greater than the average number of times being detected of all users.

3.3.2 Channel Estimation

By **Algorithm 1**, the active user index and the total number of active user are acquired, and then these information combined with time-invariant multipath delay as a prior knowledge is input to OMP. Accordingly, we can arrive at

$$\mathbf{y} = \mathbf{\Theta}_{I_C,\Delta_{I_C}} \mathbf{h}_{I_C} + \mathbf{z}, \tag{3.9}$$

where Δ_{I_C} denotes the possible multipath delay combined with detected user set I_C. Hence, CSI can be obtained by utilizing (3.9) and OMP algorithm. Table 3.1 demonstrates the computational complexity of the proposed TIDCS algorithm, and the existing OMP and CoSaMP algorithms in terms of the number of complex additions and multiplications.

Table 3.1: Complexity analysis of the proposed TIDCS, OMP and CoSaMP

Item	**Complexity**
CoSaMP [79]	$O(MNL_{\mathrm{M}})$
OMP [80]	$O(KLMNL_{\mathrm{M}})$
TIDCS	$O(KMN(2\delta+1)L^2)$

It is observed that CoSaMP has the least complexity among the three algorithms since its complexity is independent of the sparsity level, *i.e.,* KL. Hence, the complexity of

CoSaMP is selected as the benchmark. In contrast, the complexities of the proposed TIDCS and OMP depend on the exact sparsity level, and the ratio of the complexity of TIDCS over that of OMP is expressed as

$$\xi = \frac{(2\delta + 1)L}{L_M}. \tag{3.10}$$

Furthermore, the multipath delays in discrete time domain could be consecutive and small. Therefore, the multipath delays may all drop in $[1, L+\delta]$. Typically, $L = 0.1L_{\mathrm{M}}$[76]. Hence, it can be concluded that

$$\frac{L + \delta}{L_M} \leq \xi \leq \frac{(2\delta + 1)}{10}. \tag{3.11}$$

Typically, the numerator is smaller than the denominator for the upper bound since δ could be small. In fact, ξ in (3.11) ranges from 0.2 to 0.56 with the parameter refers to Section 3.3, implying the complexity of TIDCS for channel estimation is only 20 percent of OMP in the best-case scenario and nearly half of the complexity of OMP in the worst-case scenario. Furthermore, CoSaMP has lower complexity but with the worst channel estimation performance.

3.4 Simulation Results

In this section, the performance of the proposed TIDCS algorithm is presented. The total number of users N and number of subcarriers M in each symbol are both set to 256. The values of L_M and L are 26 and 3, respectively[76]. L is assumed to follow uniform distribution. The probability of active user ranges from 0.05 to 0.075. Parameter δ in (3.10) is 2. In addition, the threshold of reconfirming active user set is an integer that is no less than half of the number of valid taps. Finally, pseudo-noise is utilized to construct a Toeplitz Matrix satisfying the requirement of the restricted isometry property (RIP), which is a necessary condition for CS [47].

Particularly, the number of erroneously estimated indices should be taken into account in case that the size of recovered active user set is much greater than the real active user set, which will dramatically increase the complexity of receiver, for instance, increasing the time of responding to users in idle state. Thus, a relative error ratio (RER) is introduced, which is the deviation of the number of erroneously estimated indices from the real number

of active users. It is expressed as

$$\mathrm{RER} = \frac{1}{J}\sum_{j=1}^{J}\frac{\left|I_c^j \backslash (I_c^j \cap I)\right|}{|I|}, \tag{3.12}$$

where I represents the real active user set and J is the number of received symbols. A success rate of UAD is defined, which is the ratio between the number of accurately estimated indices and the real number of active users. A larger value of success rate of UAD means more accurate detection and vice versa. It is expressed as

$$\frac{1}{J}\sum_{j=1}^{J}\frac{\left|I_c^j \cap I\right|}{|I|}. \tag{3.13}$$

In Figure 3.3, the RER performance of the different algorithms is demonstrated under different SNR configurations. It can be seen that g-TIDCS and m-TIDCS show a lower RER performance over the benchmarks. It is because DACS algorithm only consider whether the active user is detected without an eliminating error of UAD mechanism. Furthermore, the g-TIDCS presents a higher erratic fluctuation than the m-TIDCS because the former highly subjects to that the accumulative number of being detected for different active users should be similar during iteration.

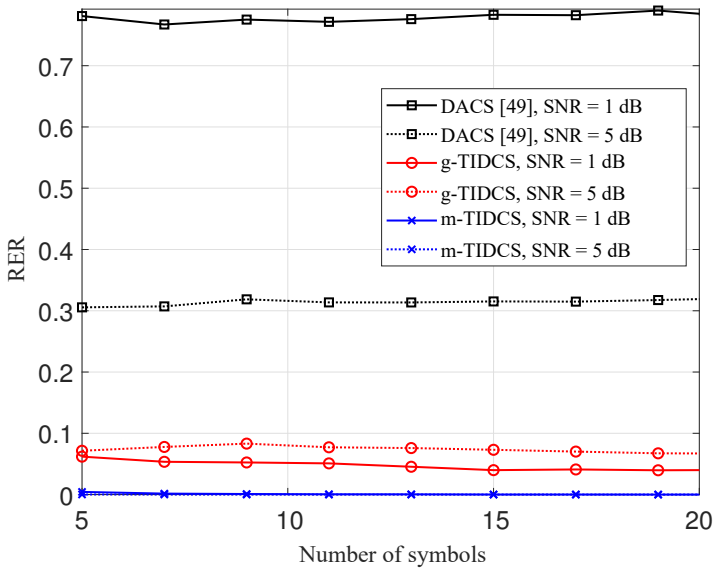

Figure 3.3: RER for the proposed TIDCS and DACS with 5% active ratio.

Given a large number of active users, *e.g.,* 7.5% active ratio, the RER performance of both m-TIDCS and g-TIDCS algorithms remains unchanged, as shown in Figure 3.4. In particular, the RER of m-TIDCS can converge to zero at SNR = 5 dB. In brief, m-TIDCS is better for low active ratio scenario from the perspective of RER.

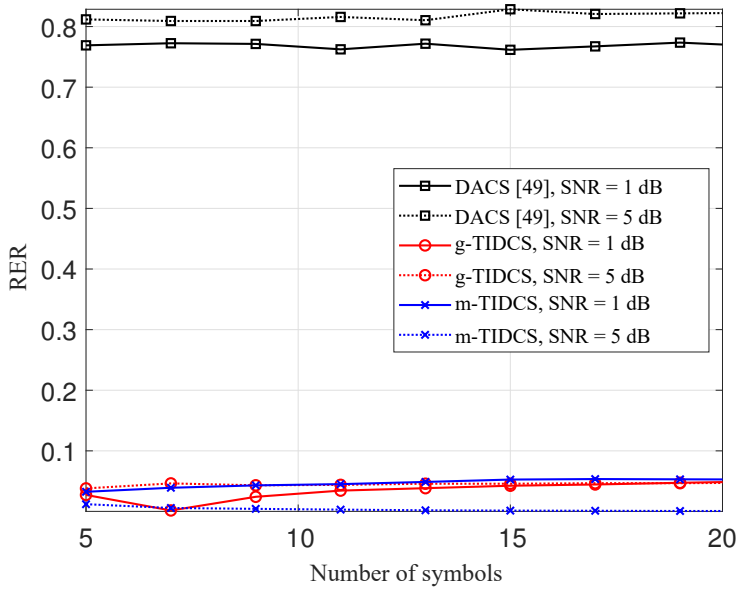

Figure 3.4: RER for the proposed TIDCS and DACS with 7.5% active ratio.

Figure 3.5 shows the impact of number of symbols on the success rate of UAD, where SNR is set to 1dB. It can be seen that m-TIDCS, g-TIDCS surpass DACS with 5% active ratio at the symbol of 13 and 19, respectively. In addition, both of proposed schemes can converge to 100% success rate of UAD. However, the performance floor appears with 7.5% active ratio for g-TIDCS due to more active users, showing that m-TIDCS is more suitable to low SNR scenario.

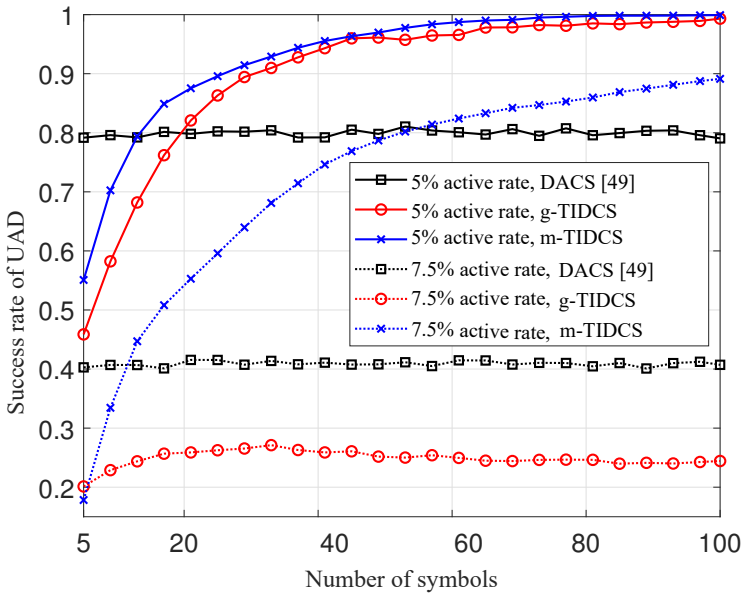

Figure 3.5: Success rate of UAD for the proposed TIDCS and DACS at SNR = 1 dB.

In Figure 3.6, where SNR is set to 5 dB, both g-TIDCS and m-TIIDCS significantly outperform DACS [49], which suffers an error floor independent of the active user ratio. At a 5% active user ratio, g-TIDCS and m-TIDCS achieve a near-perfect success rate of UAD with the number of symbols more than 40 and g-TIDCS has higher accuracy with a faster convergence speed. At a 7.5% active user ratio, m-TIDCS has a higher performance gain over g-TIDCS. Hence, m-TIDCS is more suitable for high active user ratio and very low SNR scenarios and g-TIDCS is more suitable for a higher SNR case (still in the range of low SNR).

In Figure 3.7, NMSE is demonstrated to evaluate the proposed uplink transmission scheme for CSI estimator with OMP and CoSaMP [79]. Besides, the comparison of performance only considers detected active users for the purpose of unbiased evaluation. Figure 3.7 presents that the comparison of different algorithm with 7.5% active ratio in low SNR scenario, and shows the proposed TIDCS algorithm achieves lower NMSE than the works in [79] and [80], especially for low SNR.

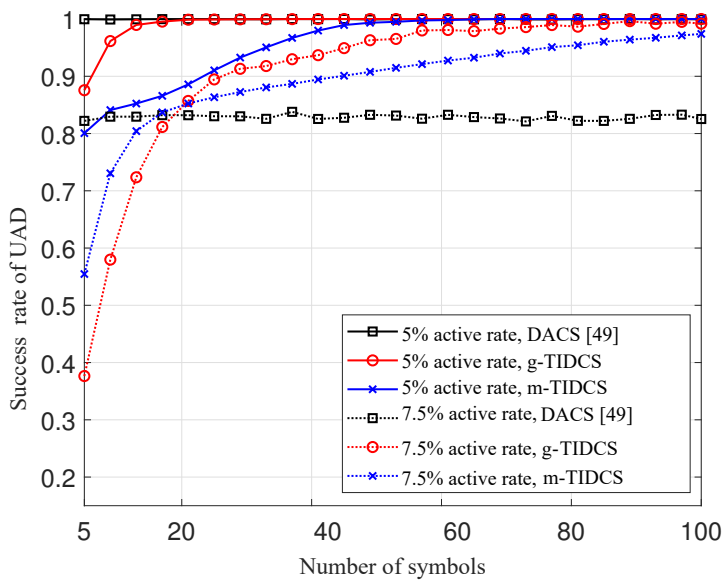

Figure 3.6: Success rate of UAD for the proposed TIDCS and DACS at SNR = 5 dB.

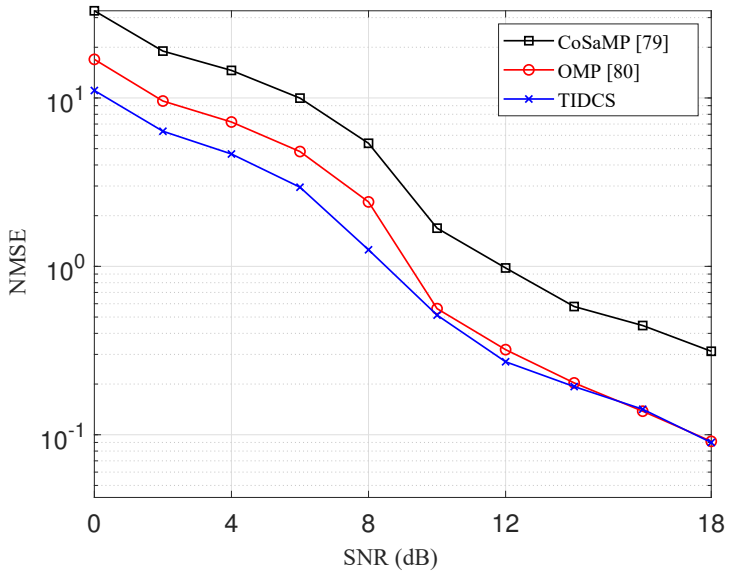

Figure 3.7: Normalized MSE performance of CSI vs. transmit SNR for proposed TIDCS algorithm, OMP and CoSaMP.

3.5 Summary

In this chapter, two CS aided UAD algorithms for grant-free communication have been presented, referred to as m-TIDCS and g-TIDCS algorithms. Taking the time-invariant channel tap delays as the prior information, the knowledge of the sparsity of users' activity is not required, which makes the algorithms more practical than the previous work [47, 52, 62, 78]. The proposed m-TIDCS and g-TIDCS algorithms achieve higher UAD accuracy than the DACS algorithm in [49] at low SNRs. Based on the obtained active users, a channel estimation scheme that has lower NMSE than CoSaMP [79] and OMP [80] is further proposed. In addition, the upper and lower bounds on the ratio of complexity between the proposed channel estimation scheme and OMP have been derived, which shows the superiority of the proposed scheme in terms of complexity.

The CS aided UAD algorithms have many advantages, such as low complexity and high success rate of UAD. Nevertheless, the proposed m-TIDCS and g-TIDCS algorithms assumed that the signals generated from different users are continuously transmitted to the BS, which can not be satisfied for some scenarios. For example, data for periodic communication is periodically collected from a large number of users and transmitted to an AP [9, 65]. Meantime, the transmission period of various users may adaptively change due to the specific requirements of energy-saving or complex control [66, 67]. In this context, the following chapter focuses on period estimation and burst user detection for grant-free communication.

Chapter 4

Adaptive Period Estimation and Burst User Detection for Periodic Grant-Free Communication

4.1 Introduction

Periodic communication is one of the salient features of IoT, where an online adjustment of the transmission period is considered to reduce the energy consumption of the sensors and actuators. Such a self-adaptive transmission period enables the controller to counteract the disturbances promptly, efficiently, and without affecting the system stability. As described in Chapter 2, the existing CS based schemes do not involve periodic communication, in which the transmission period of various users may adaptively change. This periodic communication pattern is referred to as the periodic block-sparsity structure, where the data is sparsely transmitted in a structure of the periodic block, as detailed later. Hence, intrinsic information (*e.g.*, non-continuous data temporal correlation among the periodic communication) is not considered in the existing CS based schemes.

In this chapter, joint perform CS based period estimation is investigated, UAD, channel estimation and signal detection for grant-free communication. However, the number of active users and transmission period are unknown to AP, where the exact indices of active user need to be known before performing channel estimation and signal detection. Furthermore, since the knowledge of transmission period is not known by the AP, it is difficult to

exploit the correlation between periodically transmission data. To this end, based on the non-continuous temporal correlation of the received signals and the estimated transmission period, a PBOMP algorithm is proposed. Then a PBSBL algorithm is proposed, based on the posterior distribution of the transmitted data. The major contributions of this chapter are summarized as follows:

- To the best of my knowledge, this is the first work to perform CS based channel estimation and signal detection for grant-free communication with adaptive transmission period. Due to the non-continuously received signal, the conventional CS based work [48, 49, 50, 51, 52, 53, 54, 55] that relies on the continuous temporal correlation of received signal becomes inapplicable. In contrast, the proposed design exploits the non-continuous temporal correlation of the received signal for channel estimation and detection, and therefore is more practical in uplink grant-free systems where its transmission is featured by sparsity and periodicity.

- Two algorithms for joint CS based UAD, channel estimation and signal detection, *i.e.*, PBOMP and PBSBL are proposed, without requiring the knowledge of the number of active users in the context of the periodic block-sparsity structure. By exploiting the non-continuous temporal correlation of the received data, the PBOMP and PBSBL possess the superiority of the success rate of UAD, bit error rate (BER), NMSE of channel estimation and accuracy of period estimation over their counterparts in [63] and [64].

- The CRLBS of channel estimation by the proposed PBOMP and PBSBL schemes are derived. It shows that PBSBL enjoys a lower CRLB of channel estimation than PBOMP in low and moderate SNR cases if Binary Phase Shift Keying (BPSK) or its phase-rotated variant is used for modulation. Simulation results show that PBOMP and PBSBL have close CRLBs and NMSE at high SNR.

- The proposed PBOMP and PBSBL schemes enjoy a lower complexity than their counterparts in [63] and [64]. As period estimation embedded in the proposed schemes is utilized to retrieve the potential signal directly from the received data at the AP, the number of iteration in the proposed PBOMP and PBSBL could be drastically decreased. Furthermore, the PBOMP with low complexity is more suitable for high SNR than PBSBL.

The rest of this chapter is organized as follows. System model is illustrated in Section 4.2. The proposed PBOMP and PBSBL are demonstrated in Section 4.3. The performance analysis and complexity of the proposed PBOMP and PBSBL are discussed in Section 4.4 and 4.5, respectively. Simulation results are given in Section 4.6, followed by conclusion in Section 4.7.

4.2 System Model

As shown in Figure 4.1, this chapter considers grant-free communication with the periodic block-sparsity structure. Considering the massive number of users, an overloaded uplink transmission is assumed. That is, the total number of users N is much larger than the number of the orthogonal sub-carriers M, while the number of active users K is far less than the total number of users N. In this typical IoT scenario, there are massive potential users but only a fraction of them are active for data transmission. In particular, NOMA is considered, where the active users share same frequency by code-domain for multiuser access [81]. The users periodically transmit data, where the transmission period is P. For simplicity, it is assumed that the period is unchanged in J symbols and consider a slow flat Rayleigh fading channel. The signal is received and decoded by an AP. The AP and users are equipped with single antenna. Without loss of generality, the first symbol of each block serves as the pilot signal and the remaining of the block is allocated for carrying data. Block length is denoted as Q, and the transmission period P is the multiples of Q for simplicity. In general, the aforementioned assumptions can be summarized in Table 4.1.

The code-domain multiplexing is considered, where data from different active users are spread onto different MA signatures (code) to manage inter-user interference, and the superimposed signal is transmitted over the orthogonal sub-carriers. In detail, the coded bit stream is first mapped to a modulated symbol for each user. After the symbol is modulated onto a spreading sequence S, *i.e.*, the MA signature, the active users simultaneously transmit their signals through different channels following the inverse Fast Fourier Transform (IFFT) module, and the superimposed signals received at the AP are extracted from the inverse process of IFFT. Refer to [22], a single-measurement vector based CS is used to formulate NOMA as a CS problem, where a one shot transmission is considered by taking the received signals as a vector $\boldsymbol{y}$. In this context, the superimposed signal $\mathbf{y}^j$ for the j-th

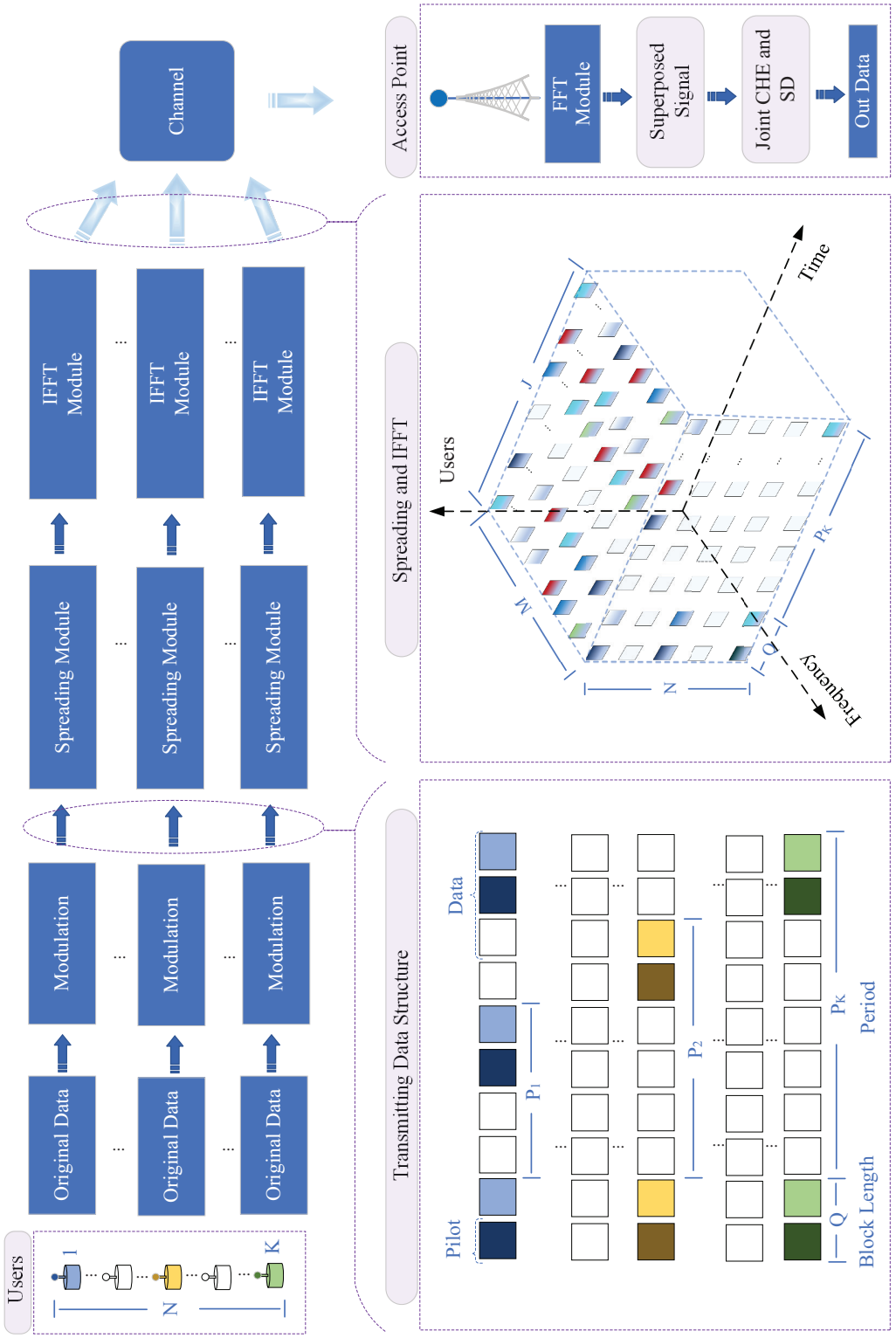

Figure 4.1: A typical communication system of IoT for periodic block-sparsity structure with N users and K active users.

Table 4.1: Main assumptions in the system model for periodic grant-free communication

Item	Constraints
An overloaded uplink transmission	$M < N$ [47, 48, 49, 82].
Sparse active users	$K \ll N$ [22, 47, 48, 49, 82].
Channel	A slow Rayleigh fading channel [47, 48, 51].
Block Structure	The first symbol serves as the pilot and the remaining of it is data [83, 74].

symbol can be modeled as[1],

$$\mathbf{y}^j = \sum_{n=1}^{N} h_n d_n^j \mathbf{s}_n + \mathbf{z}^j, \tag{4.1}$$

where d_n^j indicates transmitted signals for the j-th symbol of the n-th user and $\mathbf{s}_n$ is a spreading sequence for the n-th user. h_n denotes the fading channel between the n-th user and AP. $\mathbf{z}^j$ represents the additive white Gaussian noise for the j-th superimposed symbol at the AP. For brevity, (4.1) is collected into a compact form as

$$\mathbf{Y} = \mathbf{X}\mathbf{\Theta} + \mathbf{Z}, \tag{4.2}$$

where $\mathbf{Y} = [\mathbf{y}^1, \mathbf{y}^2, \cdots, \mathbf{y}^J]^{\mathrm{T}} \in \mathbb{C}^{J \times M}$ denotes received matrix for J superimposed symbols. $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^J]^{\mathrm{T}} \in \mathbb{C}^{J \times N}$ represents composited matrix of transmitted signals and channel, where $\mathbf{x}^j = [d_1^j h_1, d_2^j h_2, \cdots, d_n^j h_n]^{\mathrm{T}}$. $\mathbf{\Theta} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N]^{\mathrm{T}} \in \mathbb{C}^{N \times M}$ and $\mathbf{Z} = [\mathbf{z}^1, \mathbf{z}^2, \cdots, \mathbf{z}^J]^{\mathrm{T}} \in \mathbb{C}^{J \times M}$ denote spreading sequence matrix and additive white Gaussian noise (AWGN) matrix, respectively.

[1]Note that the equations in this chapter build on the frequency domain without considering the impact of IFFT and FFT.

4.3 CS based Burst User Detection for Periodic Block-Sparsity Structure

In this section, the problem formulation of joint UAD, channel estimation and signal detection with periodic block-sparsity structure in 4.3.1 is first given. Then, two algorithms, namely PBOMP and PBSBL, are introduced in 4.3.2 and 4.3.3, respectively.

4.3.1 Problem Formulation

For the periodic block-sparsity structure of grant-free communication, the exact transmitted period is not known at the AP. To detail the periodic block-sparsity structure, $\text{supp}(\mathbf{d}^j)$ as the support set is first defined, which represents the index set of nonzero elements in $\mathbf{d}$, where $\mathbf{d}^j = [d_1, d_2^j, \cdots, d_N^j]^{\text{H}}$ is the j-th transmitted symbols of the N users. Since the data block is transmitted in period P, nonzero elements in $\mathbf{d}$, *i.e.*, the common set, appear in cycle for active users. Hence, this structure can be expressed as,

$$\begin{aligned} \text{supp}(\mathbf{d}^1) = \cdots = \text{supp}(\mathbf{d}^Q) = \text{supp}(\mathbf{d}^{1+P}) \\ = \cdots = \text{supp}(\mathbf{d}^{Q+P}) = \cdots, \end{aligned} \tag{4.3}$$

where $\text{supp}(\mathbf{d}^j) = \{k : d_k^j \neq 0, 1 \leq k \leq K\}$. Refer to (4.3), it is clear that the redundant data can be eliminated if transmission period P is obtained. By doing so, the non-continuous transmission data turns into continuous transmission data, and hence the correlation between the transmission data can be sufficiently utilized.

As outlined in Section 4.2, the number of unknown variables, *i.e.*, the size of composited matrix $\mathbf{X}$, is larger than the number of equations, *i.e.*, the size of received matrix $\mathbf{Y}$. Hence, (4.2) is referred to as the underdetermined problem and the conventional methods of solving the linear equations is not applicable. In fact, the number of active users K accounts for a small fraction of the total potential users N in IoT scenario, and thus I am motivated to apply CS to jointly perform period estimation, UAD, channel estimation and signal detection. Since (4.2) gives a matrix representation of the received signals, it is extended to a vector format, such as

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{x}}\widetilde{\mathbf{\Theta}} + \widetilde{\mathbf{z}}, \tag{4.4}$$

where $\widetilde{\mathbf{y}} = \text{vec}(\mathbf{Y})$, $\widetilde{\mathbf{\Theta}} = \mathbf{\Theta}^{\text{T}} \otimes \mathbf{I}$, $\widetilde{\mathbf{x}} = \text{vec}(\mathbf{X})$, $\widetilde{\mathbf{z}} = \text{vec}(\mathbf{Z})$, and $\mathbf{I} \in \mathbb{C}^{J \times J}$ is an identity matrix.

The objective is to recover the transmitted signals $\widetilde{\mathbf{x}}$ from the superimposed signals $\widetilde{\mathbf{y}}$ without the knowledge of active users index, by exploiting the periodic block-sparsity structure defined in (4.3). Hence, the problem formulation is given as,

$$\begin{aligned} &\min_{\widetilde{\mathbf{x}}} \left\| \widetilde{\mathbf{y}} - \widetilde{\mathbf{\Theta}} \widetilde{\mathbf{x}} \right\|_2 \\ &\text{s.t. } (\text{C1}) : K < N, \\ &\quad (\text{C2}) : (4.3). \end{aligned} \tag{4.5}$$

The constraint (C1) is in line with scenarios where the number of active user K for uplink transmission is far less than the total number of users N, while the constraint (C2) is built on the fact that the data block generated by different users is transmitted in a time period denoted as P.

4.3.2 Periodic Block Orthogonal Matching Pursuit

Since the optimal solution vector $\widetilde{\mathbf{x}}$ of (4.5) is sparse, *i.e.*, only a faction of nonzero elements in $\widetilde{\mathbf{x}}$, the receiving vector $\widetilde{\mathbf{y}}$ is a linear combination of a few columns $\widetilde{\boldsymbol{\theta}}$ from $\widetilde{\mathbf{\Theta}}$. Intuitively, to find the solution of (4.5), these columns should be captured. For example, based on the correlation between the columns $\widetilde{\boldsymbol{\theta}}$ and the remaining part of $\widetilde{\mathbf{y}}$ at each iteration, one can find the columns $\widetilde{\boldsymbol{\theta}}$ in a greedy fashion, where the remaining part of $\widetilde{\mathbf{y}}$ denotes the difference between the receiving vector $\widetilde{\mathbf{y}}$ and reconstructed signal [84]. In particular, the selected columns form a subspace which is orthogonal to the remaining part of $\widetilde{\mathbf{y}}$, and the least square (LS) algorithm is applied to obtain global solution or local optimal solution. However, the classical BOMP algorithm requires the knowledge K and fails to consider the impact of periodic block-sparsity structure.

In order to find the columns of $\widetilde{\mathbf{\Theta}}$ corresponding to nonzero values of $\widetilde{\mathbf{x}}$ by exploiting the temporal correlation in the periodic block-sparsity structure. Let begin by choosing the column of $\widetilde{\mathbf{\Theta}}$ that is the most aligned with the remaining part of $\widetilde{\mathbf{y}}$, *i.e.*, the residual $\mathbf{r}^{k-1}$ between $\widetilde{\mathbf{y}}$ and the reconstruct signal at the k-th iteration, written as,

$$\lambda_k = \underset{\lambda_k, \Omega_{\lambda_k} \in \Omega}{\arg\max} \left\| \widetilde{\mathbf{\Theta}}_{\Omega_{\lambda_k}}^{\mathrm{H}} \mathbf{r}_{\Omega_{\lambda_k}}^{k-1} \right\|_2^2, \tag{4.6}$$

where $\Omega = \{\Omega_{\lambda_k}\}|_{k=1}^{N}$ denotes the user index for all users. $\Omega_{\lambda_k} = \{\Omega_{\lambda_{k,u}}\}|_{u=1}^{J}$ chooses element index that match with nonzero values of $\widetilde{\mathbf{x}}$, where $\Omega_{\lambda_{k,u}} = B(\lambda_k - 1) + u$. After identifying

the user index, the LS method is applied to make local optimal decisions [84], and the corresponding subblock index β_k can be given by

$$\beta_k = \underset{\beta_k, \Xi_{\beta_k}}{\arg\max} \left\| \widetilde{\mathbf{\Theta}}^{\dagger}_{\Xi_{\beta_k}} \widetilde{\mathbf{y}}_{\Xi_{\beta_k}} \right\|_2^2, \tag{4.7}$$

where $\Xi_{\beta_k} = \{\Xi_{\beta_k,u}\}|_{u=1}^{Q}$ is the index set of β_k-th subset of $\Omega_{\lambda_k} \backslash \widetilde{\Lambda}$ and $\Xi_{\beta_k,u} = Q(\beta_k - 1) + u$. $\Omega_{\lambda_k} \backslash \widetilde{\Lambda}$ is denotes the complement of subset in Ω_{λ_k}. $\widetilde{\Lambda}$ denotes a collection set that has been selected from the position with nonzero values of $\widetilde{\mathbf{x}}$ before the k-th iteration. To obtain the set $\widetilde{\Lambda}$, the active user index β_k is firstly merged into a user index set Γ^k, *i.e.*,

$$\Gamma^k = \Gamma^{k-1} \cup \lambda_k. \tag{4.8}$$

Then the subblock index β_k is merged into a subblock index set Λ^k, *i.e.*,

$$\Lambda^k = \Lambda^{k-1} \cup \beta_k. \tag{4.9}$$

Considering the periodic block-sparsity structure, the active user set is assumed to be unchanged in several non-continuous time slots, which leads to the temporal correlation of the received data at the AP. It motivates us to employ transmission period P to rapidly identify the possible user index. By doing so, the temporal correlation of the non-continuous transmission data can be utilized to capture the active users refer by (4.6). To this end, P is collected by finding the largest probability of occurrence of sampling space in the subblock index set Λ^k. In this context, the sampling space χ^k is defined as the set of the difference between elements of Λ^k at k-th iteration, which is iteratively calculated by

$$\xi = \Lambda^k[\mathrm{t}] - \Lambda^k[\mathrm{t}-1], \tag{4.10}$$

where $\Lambda^k[\mathrm{t}]$ denotes the t-th element of Λ^k.

Accordingly, the refined m-th subblock index of the n-th user $\Lambda_{n,m}$ of Λ can be expressed as $\Lambda_{n,m} = \{\Lambda_{n,m,\mu}\}|_{\mu=1}^{Q}$, where $\Lambda_{n,m,\mu} = J\left(\Gamma^k[n] - 1\right) + Q\left(\Lambda^k[1] - 1\right) + PQ(m-1) + \mu$. After $\Lambda_{n,m}$ is acquired, the set $\widetilde{\Lambda}$ is obtained by collecting all $\Lambda_{n,m}, \forall n, m$. Then, the reconstructed signal can be given by [84]

$$\widetilde{\mathbf{x}}_{\widetilde{\Lambda}} = \underset{\mathrm{supp}(\widetilde{\mathbf{x}}) \subset \widetilde{\Lambda}}{\arg\min} \left\| \widetilde{\mathbf{y}} - \widetilde{\mathbf{\Theta}}_{\widetilde{\Lambda}} \widetilde{\mathbf{x}}_{\widetilde{\Lambda}} \right\|_2^2. \tag{4.11}$$

Subsequently, the residual $\mathbf{r}^k$ is calculated by

$$\mathbf{r}^{k-1}=\widetilde{\mathbf{y}}-\widetilde{\mathbf{\Theta}}_{\widetilde{\Lambda}}\widetilde{\mathbf{x}}_{\widetilde{\Lambda}}. \tag{4.12}$$

Since the number of active users K is unknown by the grant-free system, a stop condition should be designed to avoid the reconstructed error caused by the idle users, *i.e.*, redundant estimation of active users. In fact, when the support set Λ^k is perfectly recovered, the power of the residual signal is equal to the power of noise, *i.e.*,

$$\left\|\widetilde{\mathbf{y}}-\widetilde{\mathbf{\Theta}}_{\widetilde{\Lambda}}\widetilde{\mathbf{x}}_{\widetilde{\Lambda}}\right\|_2^2=\|\widetilde{\mathbf{z}}\|_2^2. \tag{4.13}$$

Accordingly, when the power of residual $\mathbf{r}^k$ is less than the power of noise at the k-th iteration, the iteration of finding active users and recovering data block can be terminated. Furthermore, when the power of residual $\mathbf{r}^k$ at the k-th iteration is larger than the power of residual $\mathbf{r}^{k-1}$ at the (k-1)-th iteration, the iteration is terminated to avoid the reconstructed error. Hence, the stop condition is given by,

$$\left\|\mathbf{r}^{k-1}\right\|_2^2 \leq \left\|\mathbf{r}^{k}\right\|_2^2 \text{ or } \left\|\mathbf{r}^{k}\right\|_2^2 \leq \|\widetilde{\mathbf{z}}\|_2^2. \tag{4.14}$$

To perform signal detection, it is needed to obtain channel information. To this end, the index set of pilot in each block is firstly collected. Recall that the first position of each subblock of $\widetilde{\mathbf{x}}$ is a pilot, the pilot position set of the n-th user Ψ_n can be collected by the active user set Γ^k. In this context, Ψ_n is expressed as $\Psi_n = \{\Psi_{n,m}\}|_{m=1}^{M}$, where M denotes the total number of subblock in J symbols, and $\Psi_{n,m} = J\left(\Gamma^k[n]-1\right)+Q\left(\Lambda^k[1]-1\right)+PQ(m-1)+1$.

To obtain the channel information, it is assumed the pilot symbol for the n-th user is set to be 1, *i.e.*, $\mathbf{d}_{\Psi_n} = 1$, where $\mathbf{d} = [\mathbf{d}_1^{\mathrm{H}}, \cdots, \mathbf{d}_N^{\mathrm{H}}]^{\mathrm{H}}$. Hence the channel gain can be extracted from

$$\mathrm{h}_n = \mathbb{E}_{\widetilde{\mathbf{x}}}[\mathrm{vec}^{-1}(\widetilde{\mathbf{x}}_{\Psi_n}) * \mathbf{s}_n^{-1}]. \tag{4.15}$$

Next, the transmitted signals for the n-th user of the m-th block is calculated by

$$\mathbf{d}_{\Lambda_{n,m}} = \mathrm{vec}^{-1}(\widetilde{\mathbf{x}}_{\Lambda_{n,m}})(\mathbf{s}_n h_n)^{-1}. \tag{4.16}$$

Finally, based on the analysis above, a so called periodic BOMP algorithm is proposed, which is summarized in **Algorithm 2**. The proposed PBOMP algorithm adaptively ac-

quires the number of active users without the knowledge of K. In particular, period P is simultaneously captured and fed back to the process of identifying active users, which accelerates PBOMP convergence speed.

4.3.3 Periodic Block Sparse Bayesian Learning

In Subsection 4.3.2, PBOMP is proposed to solve (4.5), which depends on the exact reconstructed signals and the support set, *i.e.*, the selected index of active users in each iteration. The PBOMP inherits the advantage of the low complexity of the greedy algorithm, however, to exactly reconstruct signals at each iteration may be impractical especially for low SNR environments. Different from PBOMP, the periodic block-sparsity structure and SBL is considered to evaluate the posterior distribution of the unknown variables $\widetilde{x}$ conditioned on the received signals $\widetilde{y}$ and the posterior mean and variance. By doing so, the solution of (4.5) relies on the posterior distribution of the unknown variables $\widetilde{x}$ and does not need to reconstruct signals in each iteration. Specifically, the following operation is resorted to extend (4.4) to the real number field,

$$\underbrace{\begin{bmatrix} \widetilde{\mathbf{y}}_{\mathrm{R}} \\ \\ \widetilde{\mathbf{y}}_{\mathrm{I}} \end{bmatrix}}_{\widehat{\mathbf{y}}} = \underbrace{\begin{bmatrix} \widetilde{\Theta}_{\mathrm{R}} & -\widetilde{\Theta}_{\mathrm{I}} \\ \\ \widetilde{\Theta}_{\mathrm{I}} & \widetilde{\Theta}_{\mathrm{R}} \end{bmatrix}}_{\widehat{\Theta}} \underbrace{\begin{bmatrix} \widetilde{\mathbf{x}}_{\mathrm{R}} \\ \\ \widetilde{\mathbf{x}}_{\mathrm{I}} \end{bmatrix}}_{\widehat{\mathbf{x}}} + \underbrace{\begin{bmatrix} \widetilde{\mathbf{z}}_{\mathrm{R}} \\ \\ \widetilde{\mathbf{z}}_{\mathrm{I}} \end{bmatrix}}_{\widehat{\mathbf{z}}}, \tag{4.17}$$

where $(\cdot)_{\mathrm{R}}$ and $(\cdot)_{\mathrm{I}}$ extract the real part and the complex part, respectively. For brevity, $\widehat{\mathbf{y}}$, $\widehat{\mathbf{\Theta}}$, $\widehat{\mathbf{x}}$, and $\widehat{\mathbf{z}}$ are defined to simplify representation of (4.17), as shown in the underscripts. Now, a Gaussian likelihood model is considered in the context of a real number field and SBL [85, 86],

$$p\left(\widehat{\mathbf{y}} \mid \widehat{\mathbf{x}}; \sigma^2\right) = \left(2\pi\sigma^2\right)^{-MJ/2} \exp\left(\frac{-1}{2\sigma^2}||\widehat{\mathbf{y}} - \widehat{\mathbf{\Theta}}\widehat{\mathbf{x}}||^2\right), \tag{4.18}$$

where σ^2 is the noise variance. In the framework of PBSBL, $\widehat{\mathbf{x}}$ is divided into $2N$ sub-block, in which $\widehat{\mathbf{x}}_i$ is the i-th block and indicates J continuous received symbols of real part or complex part for a potential users. Furthermore, each $\widehat{\mathbf{x}}_i$ is assumed to follow a

Algorithm 2 PBOMP Algorithm

Input:

Received signals $\mathbf{Y}$;
Sensing matrix $\mathbf{\Theta}$;

Output:

Transmission period P,
activity set $\widetilde{\Lambda}$,
channel Gain $\mathbf{h}$,
transmitted data $\mathbf{d}$;

Initialization:
The number of subblock $B = J/Q$;
Initial residual $\mathbf{r}^0 = \widetilde{\mathbf{y}}$;
for $k = 1$ to N **do**
  Acquire and merge user index
  by (4.6) and (4.8);
  Acquire and merge subblock index
  by (4.7) and (4.9);
  for $t = 2$ to $\left|\Gamma^k\right|_C$ **do**
    Calculate sampling space by (4.10);
    Merge the sampling space by $\chi^k{=}\chi^k \cup \xi$;
  end for
  Extract P by $\max(\chi^k)$
  and calculate the number of subblock M by $\left\lfloor {(B-\Lambda^k[1])}/{P} \right\rfloor$;
  Refine the collected index by feedback P
  and form $\widetilde{\Lambda}{=}\left\{\{\Lambda_{n,m}\}\,|_{m=1}^{M}\right\}\Big|_{n=1}^{|\Gamma^k|_C}$;
  Reconstruct signal refer to (4.11);
  Calculate the residual signal refer to (4.12);
  If (4.14) is satisfied, then quit the iteration;
end for
Obtain the channel gain refer to (4.15), $\forall n$;
Calculate the modulated data refer to (4.16), $\forall n$.

parameterized multivariate Gaussian distribution[87], which can be written as,

$$p\left(\widehat{\mathbf{x}}_i;\beta_i,\mathbf{D}_i\right)\sim\mathcal{N}\left(\mathbf{0},\beta_i\mathbf{D}_i\right),i=1,\ldots,2N, \tag{4.19}$$

where $\beta_i = 1$ and $\beta_i = 0$ denote the active state and idle state of the i-th subblock, respectively. $\mathbf{D}_i$ indicates the correlation among $\widehat{\mathbf{x}}_i$ and shows the temporal correlation of the non-continuous transmission data. Followed by the framework of Bayesian rule, the posterior for $\widehat{\mathbf{x}}$ is given by

$$p\left(\widehat{\mathbf{x}}|\widehat{\mathbf{y}};\lambda,\{\beta_i,\mathbf{D}_i\}\,|_{i=1}^{2N}\right)\sim\mathcal{N}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right), \tag{4.20}$$

where its mean is

$$\boldsymbol{\mu}=\widehat{\mathbf{D}}\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\left(\lambda^{-1}\mathbf{I}+\widehat{\boldsymbol{\Theta}}\widehat{\mathbf{D}}\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\right)^{-1}\widehat{\mathbf{y}}, \tag{4.21}$$

and covariance is

$$\boldsymbol{\Sigma}^{-1}=\widehat{\mathbf{D}}^{-1}+\lambda\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}\widehat{\boldsymbol{\Theta}}, \tag{4.22}$$

where $\lambda={}^{1}\!/_{\sigma^2}$ and $\widehat{\mathbf{D}}=\mathrm{diag}\left(\mathbf{D}_1,\cdots,\mathbf{D}_{2N}\right)$.

Since the probability density curve of a Gaussian random variable reaches the maximum at mean value, $\widehat{\mathbf{x}}$ can be obtained by

$$\widehat{\mathbf{x}}=\boldsymbol{\mu}=\mathbb{E}_{\widehat{\mathbf{x}}}\left[p\left(\widehat{\mathbf{x}}|\widehat{\mathbf{y}};\lambda,\{\beta_i,\mathbf{D}_i\}\,|_{i=1}^{2N}\right)\right], \tag{4.23}$$

when λ, β_i and correlation matrix $\widehat{\mathbf{D}}$ are specified. These unknow parameters could be estimated by the marginalized probability density function, written as,

$$\begin{aligned}p(\widehat{\mathbf{y}}\mid\boldsymbol{\beta},\lambda)&=\int_{\widehat{\mathbf{x}}}p(\widehat{\mathbf{y}}\mid\widehat{\mathbf{x}};\lambda)p(\widehat{\mathbf{x}};\widehat{\mathbf{D}})d\widehat{\mathbf{x}}\\&\sim\mathcal{N}(\mathbf{0},\mathbf{W}),\end{aligned} \tag{4.24}$$

with $\mathbf{W}=\lambda^{-1}\mathbf{I}+\widehat{\boldsymbol{\Theta}}\widehat{\mathbf{D}}\widehat{\boldsymbol{\Theta}}^{\mathrm{T}}$ and $\boldsymbol{\beta}=[\beta_1,\beta_2,\cdots,\beta_{2N}]^{\mathrm{T}}$, which is also referred to as type-II maximum likelihood [88]. Subsequently, the equivalent cost function can be expressed as,

$$\begin{aligned}\mathcal{L}\left(\{\beta_i,\mathbf{D}_i\}\,|_{i=1}^{2N},\lambda\right)&\triangleq-2\log p\left(\widehat{\mathbf{y}}\mid\{\beta_i,\mathbf{D}_i\}\,|_{i=1}^{2N},\lambda\right)\\&=\log|\mathbf{W}|+\widehat{\mathbf{y}}^{\mathrm{T}}\mathbf{W}^{-1}\widehat{\mathbf{y}}.\end{aligned} \tag{4.25}$$

Since (4.25) cannot be solved in a closed form, the expectation-maximization algorithm

(EM) is employed to derive λ, β_i and $\mathbf{D}_i$ [89, 90]. However, the performance of parameter estimation by the EM will be significantly affected by redundant noise subblock. Hence, it is required to rapidly eliminate noise subblock and identify sparse periodic block. To this end, the coefficient submatrix $\mathbf{V}_{i,\Xi_b,\Xi_b}$ of $\mathbf{D}_i$ is firstly extracted, whose elements comprise the intersection between the rows Ξ_b of $\mathbf{D}_i$ and the columns Ξ_b of $\mathbf{D}_i$, *i.e.*,

$$\mathbf{V}_{i,\Xi_b,\Xi_b} = \mathbf{D}_i[\Xi_b; \Xi_b], \tag{4.26}$$

where $\Xi_b = \{Q\,(b-1)+u\}|_{u=1}^{Q}$ is the index set of b-th submatrix of i-th coefficient matrix $\mathbf{D}_i$ and $b \in [1, B]$. Apparently,

$$||\bar{\alpha}_{n,m}||_2^2 > ||\dot{\alpha}_{n,m}||_2^2,\ n \neq m \text{ and } 1 \leq n, m \leq Q, \tag{4.27}$$

where $\bar{\alpha}_{n,m}$ denotes the element at the n-th row and m-th column of $\mathbf{V}_{i,\Xi_b,\Xi_b}$ of the sparse periodic subblock. Similarly, $\dot{\alpha}_{n,m}$ denotes the element of $\mathbf{V}_{i,\Xi_b,\Xi_b}$ of the noise subblock. Since the n-th and m-th symbols are independent from the noise subblock, $||\dot{\alpha}_{n,m}||_2^2$ approaches zero. Hence, the index set $\bar{\Xi}_b$ of $\mathbf{V}_i$ should satisfy

$$\sum_{\bar{\Xi}_b} \left\| \mathbf{V}_{i,\bar{\Xi}_b,\bar{\Xi}_b} \right\|_2^2 > \vartheta,\ \forall b, \tag{4.28}$$

where ϑ is a threshold to capture sparse periodic subblock. The transmission period P is captured by all the collected b in line with the index set $\bar{\Xi}_b$, $\forall b$. Then the period is fed back to identify the sparse periodic subblock and noise subblock, *i.e.*, to exclude the subblock without temporal correlation from the coefficient matrix $\widehat{\mathbf{D}}$. Refer to (4.28) and apply the EM method, (4.21) and (4.22) can be iteratively updated following the rules as follows,

$$\beta_i = \frac{1}{d_i} \mathrm{Tr}\left[\mathbf{D}_{i,\bar{\Xi}_b,\bar{\Xi}_b}^{-1} \left(\mathbf{\Sigma}_{i,\bar{\Xi}_b,\bar{\Xi}_b} + \boldsymbol{\mu}_{i,\bar{\Xi}_b} \boldsymbol{\mu}_{i,\bar{\Xi}_b}^{\mathrm{T}} \right) \right], \forall i, \tag{4.29}$$

$$\mathbf{D}_{i,\bar{\Xi}_b,\bar{\Xi}_b} = \frac{\left[\mathbf{\Sigma}_{i,\bar{\Xi}_b,\bar{\Xi}_b} + \boldsymbol{\mu}_{i,\bar{\Xi}_b} \boldsymbol{\mu}_{i,\bar{\Xi}_b}^{\mathrm{T}} \right]}{\beta_i},\ \forall i, \tag{4.30}$$

$$\lambda^{-1} = \frac{\mathrm{Tr}\left[\mathbf{\Sigma}_{\Xi,\Xi} \widehat{\mathbf{\Theta}}_{\Xi,\Xi}^{\mathrm{T}} \widehat{\mathbf{\Theta}}_{\Xi,\Xi} \right]}{2N} + \frac{\left\| \widehat{\mathbf{y}}_{\Xi} - \widehat{\mathbf{\Theta}}_{\Xi,\Xi} \boldsymbol{\mu}_{\Xi} \right\|}{2N}, \tag{4.31}$$

where $\boldsymbol{\mu}_i \subset \mathbb{R}^{J\times 1}$ and $\mathbf{\Sigma}_i \subset \mathbb{R}^{J\times J}$ are the i-th subblocks of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, respectively. Ξ

denotes the collections of $\bar{\Xi}_{\mathrm{b}}$ for $2N$ blocks in consistent with (4.28) and the identified transmission period P.

Based on the analysis above, the optimal solution of (4.17) can be found by iteratively calculating (4.21), (4.22), (4.28), (4.29), (4.30) and (4.31). Hence, these iterative calculations are named as period block SBL, which is summarized in **Algorithm 3**. The proposed PBSBL consists of three parts, namely Parameter Learning, Pruning with Period Estimation and Stop Iteration. Parameter Learning aims to learn parameters (*e.g.*, β_i and λ), to serve aforementioned iterative calculation. For the step of Pruning with Period Estimation, the transmission period P is extracted from all the collected b satisfying (4.28). Subsequently, P is fed back to collect Ξ and eliminate the noise subblock. Similar to (4.14) in **Algorithm 2**, Stop Iteration shows that the iteration is stopped when the difference between the k-th iteration and the (k-1)-th iteration of $\boldsymbol{\mu}$ below a critical value. After applying the the inverse process of (4.17), (4.15) and (4.16) are implemented to execute decoding after the noise subblock is excluded from $\widehat{\mathbf{x}}$.

Algorithm 3 PBSBL Algorithm

1: **for** $n = 1$ to N **do**
2: **Parameter Learning**:
3: Update $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ refer to (4.21) and (4.22);
4: $\forall i$, update subblock coefficient β_i, coefficient matrix $\mathbf{D}_i$ refer to (4.29) and (4.30); Update λ refer to (4.31);
5: **Prune** with **Period Estimation**:
6: Extract Periodic Information by (4.26), (4.27) and (4.28);
7: Eliminate the noise subblock by collecting Ξ;
8: **Stop Iteration**:
9: The difference between the current and the last $\boldsymbol{\mu}$ is below a critical value;
10: **end for**
11: Substitute the real number field to the complex field, refer to (4.17);
12: Line 17-18 in **Algorithm 2**.

4.4 Performance Analysis

In Section 4.3, two approaches, namely PBOMP and PBSBL, have been proposed to perform joint channel estimation and signal detection for uplink grant-free IoT in the context of block periodic block-sparsity structure. In this section, CRLB of channel estimation for the proposed approaches is analyzed.

4.4.1 FIM of Channel Estimation for the Proposed PBOMP and PBSBL

As a precursor to the subsection that follow, *Lemma 4.4.1* is introduced.

. Lemma 1 *(Fisher Information Matrix (FIM), [91]): Let* $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_N]$ *be a random vector and* $p(\boldsymbol{\theta})$ *be a probability on* $\boldsymbol{\theta}$ *with continuous first and second order partial derivatives. The FIM of* $\boldsymbol{\theta}$*, written as* $\mathbf{J}(\boldsymbol{\theta})$*, is given by*

$$\mathbf{J}(\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathrm{T}}], \tag{4.32}$$

where $\boldsymbol{\Delta}=\frac{\partial log p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$.

In PBOMP, $\boldsymbol{\theta}$ is equal to $\boldsymbol{h}$, and it can be obtained that

$$\log p(\widetilde{\mathbf{y}}; \mathbf{h}) = -\frac{MJ}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}||\widetilde{\mathbf{y}}-\widetilde{\boldsymbol{\Theta}}_{\widetilde{S}}\boldsymbol{h}||_2^2, \tag{4.33}$$

where $\widetilde{S}$ denotes the index set of the pilot position of the active user. Based on (4.32) and (4.33), the FIM associated with this estimation problem is

$$\mathbf{J}_{\mathrm{PBOMP}}(\mathbf{h})=\frac{-1}{\delta^2}\widetilde{\boldsymbol{\Theta}}^{\mathrm{H}}_{\widetilde{S}[i]}\widetilde{\boldsymbol{\Theta}}_{\widetilde{S}[j]},\ 1 \le i, j \le K, \tag{4.34}$$

In fact, (4.34) fails to provide any information about the previous parameter $\boldsymbol{\beta}$ defined in PBSBL. To bridge this gap, $\boldsymbol{\theta}$ is split into two groups, *i.e.*, $\boldsymbol{\theta}=\left[\boldsymbol{\theta}_{\mathrm{r}}^{\mathrm{T}}; \boldsymbol{\theta}_{\mathrm{d}}^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\boldsymbol{\theta}_{\mathrm{r}}$ attributes to deterministic parameters and $\boldsymbol{\theta}_{\mathrm{d}}$ denotes random parameters distributed according to a known PDF. To find the likelihood function involves parameter $\boldsymbol{\beta}$, $p(\mathbf{y}; \boldsymbol{\theta})$ is rewritten as $p(\mathbf{y}, \boldsymbol{\theta}_{\mathrm{r}}; \boldsymbol{\theta}_{\mathrm{d}})$, which denotes a function of the observations $\mathbf{y}$ with respect to $\boldsymbol{\theta}_{\mathrm{r}}$ and $\boldsymbol{\theta}_{\mathrm{d}}$. Hence, in PBSBL it can be concluded that,

$$p(\mathbf{y}, \boldsymbol{\theta}_{\mathrm{r}}; \boldsymbol{\theta}_{\mathrm{d}})=p(\widehat{\mathbf{y}}, \widehat{\mathbf{x}}; \boldsymbol{\beta}), \tag{4.35}$$

$$\begin{aligned}\log p(\widehat{\mathbf{y}}, \widehat{\mathbf{x}}; \boldsymbol{\beta}) = &-\frac{1}{2}\left(2MJ\log 2\pi\sigma^2 + 2N\log 2\pi + \log|\boldsymbol{\Upsilon}|\right) \\ &-\frac{1}{2}\Big(\frac{||\widehat{\mathbf{y}}-\widehat{\boldsymbol{\Theta}}\widehat{\mathbf{x}}||_2^2}{\sigma^2} + \widehat{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\Upsilon}^{-1}\widehat{\mathbf{x}}\Big),\end{aligned} \tag{4.36}$$

where $\boldsymbol{\Upsilon} = \mathrm{diag}(\beta_1\mathbf{D}_1, \cdots, \beta_{2N}\mathbf{D}_{2N})$ represents covariance matrix. Refer to (4.32), the element of the i-th and j-th of the FIM for random parameter $\widehat{\mathbf{x}}$, *i.e.*, $\mathbf{J}\left(\widehat{\mathbf{x}}\right)$, can be

computed as [91]

$$\mathbf{J}_{i,j}(\widehat{\mathbf{x}})=\begin{cases} -\frac{\widehat{\mathbf{\Theta}}_{\widehat{S}_i}^{\mathrm{T}}\widehat{\mathbf{\Theta}}_{\widehat{S}_j}}{\sigma^2}-2(\beta_i\mathbf{D}_i)^{-1}, & \text{if } i=j, \\ 0, & \text{others}, \end{cases} \tag{4.37}$$

where $\widehat{S}_i$ denotes the index set of i-th subblock in $\widehat{\mathbf{x}}$. Recall that the first symbol of each transmitting block serves as pilot and the pilot data is set to be 1. Hence, the real or imaginary part of channel information of FIM for the i-th block could be given by

$$\mathbf{J}_{\mathrm{PBSBL}}(\mathbf{h}_i)=\mathbf{J}_{i,i}(\widehat{\mathbf{x}}_{\widehat{S}_i[1]}). \tag{4.38}$$

4.4.2 CRLB Difference Between the Proposed PBOMP and PBSBL

To provide a more intuitive understanding of the difference between the proposed approaches, CRLB is employed to compare the proposed PBOMP and PBSBL, which is used in estimation theory to provide a lower limit on the variance. To this end, *Lemma 4.4.2* is introduced.

. Lemma 2 *(Cramer-Rao lower bound (CRLB), [91]): Assuming the FIM is non-singular, the CRLB matrix of the proposed approaches can be obtained by the inverse of the FIM, i.e.,*

$$var\{\boldsymbol{\theta}\} \geq (-\mathbf{J}(\boldsymbol{\theta}))^{-1}, \tag{4.39}$$

where $var\{\boldsymbol{\theta}\}$ denotes the variances of parameter $\boldsymbol{\theta}$ estimation.

Based on (4.19), the $\mathbf{D}_i$ indicates the correlation among $\widehat{\mathbf{x}}_i$. Hence, $\mathbf{D}_i$ is an identity matrix for the active state of the i-th subblock. Furthermore, the active state of the i-th subblock is in line with $\beta_i = 1$. In this context, it can be concluded that

$$\mathbf{J}_{\mathrm{PBSBL}}(\mathbf{h}_i) = -\frac{\widehat{\mathbf{\Theta}}_{\widehat{S}_i[1]}^{\mathrm{T}}\widehat{\mathbf{\Theta}}_{\widehat{S}_i[1]}}{\sigma^2}-2. \tag{4.40}$$

Refer to (4.17), $\widehat{\mathbf{\Theta}}$ is symmetric, which comprises of a real part $\widetilde{\mathbf{\Theta}}_{\mathrm{R}}$ and an imaginary part $\widetilde{\mathbf{\Theta}}_{\mathrm{I}}$ of $\widetilde{\mathbf{\Theta}}$. Hence, the item $\widetilde{\mathbf{\Theta}}_{\widetilde{S}_i}^{\mathrm{H}}\widetilde{\mathbf{\Theta}}_{\widetilde{S}_i}$ in (4.34) is equal to the item $\widehat{\mathbf{\Theta}}_{\widehat{S}_j[1]}^{\mathrm{T}}\widehat{\mathbf{\Theta}}_{\widehat{S}_j[1]}$ in

(4.40) for the same element of $\mathbf{h}$, *i.e.*, $\widetilde{S}[i] = \widehat{S}_j[1]$. In this context, after implementing (4.34) and (4.40), it is concluded that $\mathbf{J}_{\text{PBOMP}}(\mathbf{h}) > \mathbf{J}_{\text{PBSBL}}(\mathbf{h}_{\text{R}})$ or $\mathbf{J}_{\text{PBOMP}}(\mathbf{h}) > \mathbf{J}_{\text{PBSBL}}(\mathbf{h}_{\text{I}})$. Accordingly, for the real part or imaginary part of the channel information, the CRLB of channel estimation for PBSBL is lower than that of the channel estimation for PBOMP. Furthermore, it is assume to the power of the MA signature $\widehat{\mathbf{\Theta}}^{\text{T}}_{\widehat{S}_i[1]}\widehat{\mathbf{\Theta}}_{\widehat{S}_i[1]} = \omega$, it can be concluded that [2],

$$\text{var}_{\text{PBOMP}}(\mathbf{h}) - \text{var}_{\text{PBSBL}}(\mathbf{h}) = \frac{\sigma^2(2\sigma^2 - \omega)}{\omega(2\sigma^2 + \omega)}. \tag{4.41}$$

Hence, the CRLB difference of the channel estimation between the proposed PBOMP can be summarized into two parts, which can be written as,

$$\text{var}_{\text{PBOMP}}(\mathbf{h}) - \text{var}_{\text{PBSBL}}(\mathbf{h}) \begin{cases} \geq 0, \text{ if } \sigma^2 \geq \omega/2, \\ \\ < 0, \text{ if } \sigma^2 < \omega/2. \end{cases} \tag{4.42}$$

When the noise power σ^2 is greater than half of ω, it is clear that the CRLB of the channel estimation for the PBOMP is larger than or equal to the PBSBL. Furthermore, $\frac{\omega}{\sigma^2}$ approximates 0 with high noise environment, it can be concluded that

$$\text{var}_{\text{PBOMP}}(\mathbf{h}) - \text{var}_{\text{PBSBL}}(\mathbf{h}) = \frac{\sigma^2}{\omega}. \tag{4.43}$$

It could be inferred therefore that the CRLB difference is proportionate to the increased power of noise. This suggests that the performance of channel estimation for the PBSBL is superior to the PBOMP, especially for high noise scenarios. On the other hand, the PBOMP enjoys a lower CRLB of the channel estimation than the PBSBL when the noise power σ^2 is greater than half of ω. Refer to (4.41), since $\frac{\sigma^2}{\omega}$ approximates 0 with small noise power, it should, however, be noted that the CRLB of the channel estimation for the PBOMP is similar to the PBSBL. Empirically, it seems that the performance of channel estimation for the PBSBL can be similar to the PBOMP with a small noise environment.

[2]Note that $\widehat{\mathbf{\Theta}}$ is symmetric, and hence $\mathbf{J}_{\text{PBSBL}}(\mathbf{h}_{\text{R}}) = \mathbf{J}_{\text{PBSBL}}(\mathbf{h}_{\text{I}})$ refer to (4.40).

Furthermore, it is assumed that the imaginary part $\widetilde{\mathbf{\Theta}}_{\mathrm{I}}$ of $\widetilde{\mathbf{\Theta}}$ is omitted, (4.41) can be rewritten as

$$\mathrm{var}_{\mathrm{PBOMP}}(\mathbf{h}) - \mathrm{var}_{\mathrm{PBSBL}}(\mathbf{h}) = \frac{2\sigma^4}{\omega(2\sigma^2+\omega)}. \tag{4.44}$$

Hence, it can be concluded that the CRLB of the channel estimation of the PBSBL is superior to the PBOMP on the whole noise region if BPSK or its phase-rotated version is adopted as a modulation scheme.

4.5 Complexity Analysis

In this section, the computational complexity of the proposed approaches at each iteration is firstly presented. Then the comparison between the proposed approaches and other algorithms is discussed.

4.5.1 Complexity of the proposed PBOMP and PBSBL

To present a more detailed insight into different approaches in terms of computational complexity, this chapter only focuses on the number of multiplications in the theoretical analysis since multiplication plays a key role in the elapsed time required for running programs. The computational complexity of the proposed PBOMP and PBSBL are discussed in this subsection, respectively.

The main calculation of PBOMP attributes to the two processes, refer to Support Estimation in line 4 and Signal Reconstruction in line 13 of **Algorithm 2**. At each iteration, the former is dominated by the complex multiplication of the correlation operation with the complexity

$$C^1_{\mathrm{PBOMP}} = NMJ^2. \tag{4.45}$$

The latter is contributed by Least Square operation with the complexity

$$C^2_{\mathrm{PBOMP}} = (nQ)^3 + 2(nQ)^2MJ + nQMJ, \tag{4.46}$$

where n is the number of the selected data block transmitted by active users and the maximum n is KB. Hence, the computational complexity of the proposed PBOMP at

each iteration can be computed as,

$$\begin{aligned} C_{\text{PBOMP}} =& C^1_{\text{PBOMP}} + C^2_{\text{PBOMP}} \\ =& (nQ)^3 + MJ\left(2(nQ)^2 + NJ + nQ\right). \end{aligned} \tag{4.47}$$

Compared to PBOMP, the SBL based PBSBL shows a different internal structure in calculation, leading to an increasing computational complexity. As shown previously, the convergence rate of the PBSBL relies on the accuracy and efficiency of parameter learning, including mean (4.21), and variance (4.22), and hence Parameter Learning for computational complexity is considered. At each iteration, the complexity of mean $\boldsymbol{\mu}$ is given by

$$\begin{aligned} C^1_{\text{PBSBL}} =& N_k\left((MJ)^3 + M^2J^3 + MJ^3\right) \\ &+ 2N_kMJ^2 + N_kB_kQ^2, \end{aligned} \tag{4.48}$$

and variance is given by

$$C^2_{\text{PBSBL}} = (N_kJ)^3 + N_kM^2J^3, \tag{4.49}$$

where N_k is the remaining users and B_k is the remaining subblock to be detected at the k-th iteration. The maximum of N_k and B_k are $2N$ and B, respectively. Similarly, the parameter such as β, $\mathbf{D}$ and λ are respectively calculated as

$$C^3_{\text{PBSBL}} = N_kB_kQ^2\left(2Q+1\right), \tag{4.50}$$

$$C^4_{\text{PBSBL}} = N_kB_k\left(Q^2+1\right), \tag{4.51}$$

$$C^5_{\text{PBSBL}} = MN_kJ^2 + (MJ)^2 + (N_kJ)^3 + MN_k^2J^3. \tag{4.52}$$

Hence, the computational complexity of the proposed PBSBL at each iteration can be computed as,

$$\begin{aligned} C_{\text{PBSBL}} =& C^1_{\text{PBSBL}} + C^2_{\text{PBSBL}} + C^3_{\text{PBSBL}} + C^4_{\text{PBSBL}} + C^5_{\text{PBSBL}} \\ =& N_kMJ^2\left(J(M+1)^2 + N_kJ^2 + 3\right) \\ &+ N_kB_k(2Q^3 + 3Q^2 + 1) + (MJ)^2. \end{aligned} \tag{4.53}$$

To intuitively present the complexity comparison between the proposed approaches, as an example, the numerical results of the number of maximum multiplications for (4.47)

and (4.53) are shown in Table 4.2. It shows that the complexity of PBOMP is far less than that of PBSBL at each iteration based on the same parameters defined in Section 4.2.

Table 4.2: Numerical results of complexity for the proposed PBOMP and PBSBL at each iteration.

Algorithm	Number of maximum multiplications[1]
PBOMP[2]	2.2907e+09
PBSBL	1.5570e+14

[1] As an example, N, M, Q, P, J and K is set to be 256, 128, 4, 8, 40 and 13, respectively.
[2] One complex multiplication is equal to four times of real multiplication.

4.5.2 Complexity Comparison

Subsection 4.5.1 presents the computational complexity of the proposed approaches at each iteration. However, their complexity depends on their iteration number. Clearly, the fewer the number of iterations, the higher the computational efficiency of the proposed PBOMP and PBSBL. Hence, iteration number is applied to compare the proposed approaches with other algorithms in this subsection. It is noted that all analytical procedures derive from the premise that the sparse degree, *i.e.*, active users K, is known for the fairness concerns. For simplicity, it is assumed that each successfully detected subblock is successfully detected within one iteration for comparison schemes.

Since the iteration number of the proposed PBOMP and PBSBL depends on the performance of period estimation. Hence, the parameter τ is firstly defined to show the efficiency of estimating period P, a larger τ indicates the collection of P needs more iterations and vice versa. Then τ is resorted to provide an intuitive way to explain the superiority of the proposed approaches in terms of computational complexity. It is straightforward that τ is set to be one regardless of period estimation. In contrast, for the best scenario where P is obtained with minimal iteration, τ represents the ratio of the minimum iteration over the maximum iteration in the PBOMP or PBSBL. For the PBOMP, the lower boundary

of τ_{PBOMP} is parallel to that the period P and the first active user are simultaneously captured. Based on the above discussion, it can be concluded that

$$\tfrac{P}{J} \leq \tau_{\text{PBSBL}} \leq 1. \tag{4.54}$$

The upper boundary of τ_{PBOMP} represents a class of block greedy algorithms without considering the periodic sparsity-block structure, such as BOMP, BCoSaMP, block subspace pursuit (Block-SP), etc. On the other hand, the lower boundary of τ_{PBOMP} could approaches zeros, when the value of J is large. In other words, compared to the block greedy algorithm mentioned above, the number of the iteration for the PBOMP could be trivial along with the increase of the length of the received symbols.

For the PBSBL, its minimum and maximum iteration numbers are $2J(N-K)/P$ and $2J(NP-KQ)/QP$, respectively. Build on the above discussion, it can be concluded that,

$$\tfrac{(N-K)Q}{NP-KQ} \leq \tau_{\text{PBSBL}} \leq 1. \tag{4.55}$$

Particularly, the number of active users K accounts for a small fraction of the total potential users N and the lower bound can be expressed as

$$\tau_{\text{PBSBL-LB}} = \frac{\mathrm{Q}}{\mathrm{P}}. \tag{4.56}$$

Refer to the definition of τ_{PBSBL}, the maximum iteration number can be considered as the iteration number of other SBL based algorithm. Hence, the number of iterations required by the proposed PBSBL is far less than the other SBL based solutions.

The complexity analysis is summarized in Table 4.3, which consists of the proposed approaches and their counterparts (*e.g.*, BCoSaMP and ICASBL), in terms of the number of iterations for the above-mentioned steps. Refer to [63], the complexity of BCoSaMP is $\big(8(KJQ/P)^3+8(KJQ/P)^2MJ+KMQJ^2/P\big)$, which is larger than the PBOMP. As for the ICASBL[3], its computational complexity at each iteration is similar to the PBSBL as both of the ICASBL and the PBSBL are based on the Parameter Learning. Meanwhile, the Parameter Learning is a linear function of a series of variables, such as N_k, M and J, in which 3 is the highest power of these variables [64]. Accordingly, the PBSBL enjoys less computational than that of the ICASBL since the PBSBL aided by period estimation

[3]Note that this chapter mainly considers the key process of ICASBL consists of (23), (24), (29), (30) and (31) in [64].

has less iteration number. As discussed earlier, the number of iterations for the proposed PBOMP and PBSBL could both be significantly reduced, especially for the scenario where the transmission period is rapidly captured.

Table 4.3: Number of iterations for PBOMP, PBSBL, BCoSaMP and ICASBL.

Algorithm	Number of iterations
Proposed PBOMP[1]	Min: K Max: KJ/P
BCoSaMP [63][1]	KJ/P
Proposed PBSBL[2]	Min: $2J(N-K)/P$ Max: $2J(NP-KQ)/QP$
ICASBL [64][2]	$2J(NP-KQ)/QP$

[1] Classified as block Greedy Algorithm and calculated in the complex number field.
[2] Classified as SBL Algorithm and calculated in the real number field.

4.6 Simulation Results

In this section, the performance of the proposed approaches is demonstrated. Also, the impact of the coefficient matrix on the period estimation of the proposed PBSBL is discussed. The number of users N and the length of spreading sequence M are set to be 20 and 40, respectively. In particular, K accounts 10% of the total number of potential users in Figure 4.2, Figure 4.3, Figure 4.4, Figure 4.5 and Figure 4.6, while the number of active user is increased in Figure 4.7 and Figure 4.8 to evaluate the performance of

UAD and period estimation in terms of the different number of active users. The length of transmitted block Q is 4, and period P is an integral multiple of Q, *i.e.*, $2 < P/Q < 20$. ϑ in (4.28) is set to be the mean of off-diagonal elements of correlation matrix $\widehat{\mathbf{D}}$. To fulfill the requirement of the restricted isometry property, the spreading sequence is generated by the identically and independently distributed Bernoulli random variable [53]. Furthermore, BPSK is considered and four metrics are applied to evaluate the accuracy of signal reconstruction, consisting of the success rate of UAD, BER, NMSE of channel estimation and accuracy of period estimation. The success rate of UAD shows the ratio between the number of active users detected and the number of active users. Similarly, the detection accuracy of period estimation indicates the ratio between the number of correct periods of detected active users and the number of detected active users. Furthermore, BER denotes the percentage of bits that have errors relative to the total number of bits received in a transmission [92].

The two closely-related algorithms are selected as benchmarks of the simulation, *i.e.*, BCoSaMP and ICASBL. The former based on BOMP generally shows a high performance of signal reconstruction among the typical block greedy algorithms (*e.g.*, Block-SP) [63], while the latter is the extension of the typical SBL by decomposing the solution matrix into a multiplication of a sparse matrix and a linear mixing matrix [64].

Figure 4.2 provides the performance of success rate of UAD for the different algorithms. It appears that the performance of the proposed PBSBL reaches perfect detection of UAD if SNR is great than 8 dB, which is 5% up over the PBOMP and BCoSaMP at low SNR in terms of success rate of UAD. In particular, the counterpart of the proposed PBSBL, *i.e.*, ICASBL reaches perfect detection of UAD at SNR = 16 dB. Furthermore, it is also observed that the proposed approaches have the ability to reach 90% of the success rate of UAD if SNR is great than 8 dB. Hence the ability to recognize potentially active users in the proposed PBSBL and PBOMP has strong robustness for noise, which is the key element to design a successful uplink grant-free system. On the other hand, the UAD performance of PBOMP is superior to BCoSaMP. The reason behind is that the typical block greedy algorithms, such as BOMP [93], block generalized orthogonal matching pursuit[63] and BCoSaMP [63], need to capture subblock in proper order, which depend on the accurate estimation of user set at each iteration. To illustrate that, it is assumed that $\mathcal{P}_{\Lambda^k}=\widetilde{\mathbf{\Theta}}_{\Lambda^k}(\widetilde{\mathbf{\Theta}}_{\Lambda^k}^{\mathrm{H}}\widetilde{\mathbf{\Theta}}_{\Lambda^k})^{-1}\widetilde{\mathbf{\Theta}}_{\Lambda^k}^{\mathrm{H}}$ and $\mathcal{P}_{\Lambda^k}^{\perp}=\mathrm{I}-\mathcal{P}_{\Lambda^k}$ show the projector and the orthogonal complement projector on the subspace spanned by the columns of $\widetilde{\mathbf{\Theta}}_{\Lambda^k}$ with the support set Λ^k at the k-th iteration of BOMP, block generalized orthogonal matching

pursuit and BCoSaMP. The received signals at the k-th iteration can be decomposed into two components as follows,

$$\begin{aligned}\widetilde{\mathbf{y}} =&\widetilde{\mathbf{\Theta}}_{\Lambda^k}\left(\widetilde{\mathbf{x}}_{\Lambda^k}+\left(\widetilde{\mathbf{\Theta}}_{\Lambda^k}^{\mathrm{H}}\widetilde{\mathbf{\Theta}}_{\Lambda^k}\right)^{-1}\widetilde{\mathbf{\Theta}}_{\Lambda^k}^{\mathrm{H}}\widetilde{\mathbf{z}}\right)\\&+\widetilde{\mathbf{\Theta}}_{\Lambda/\Lambda^k}\widetilde{\mathbf{x}}_{\Lambda/\Lambda^k}+\mathcal{P}_{\Lambda^k}^{\perp}\widetilde{\mathbf{z}}\\\triangleq&\widetilde{\mathbf{\Theta}}_{\Lambda^k}\bar{\mathbf{x}}+\bar{\mathbf{z}},\end{aligned}\tag{4.57}$$

where $\bar{\mathbf{x}}=\widetilde{\mathbf{x}}_{\Lambda^k}+(\widetilde{\mathbf{\Theta}}_{\Lambda^k}^{\mathrm{H}}\widetilde{\mathbf{\Theta}}_{\Lambda^k})^{-1}\widetilde{\mathbf{\Theta}}_{\Lambda^k}^{\mathrm{H}}\widetilde{\mathbf{z}}$ and $\bar{\mathbf{z}}=\widetilde{\mathbf{\Theta}}_{\Lambda/\Lambda^k}\widetilde{\mathbf{x}}_{\Lambda/\Lambda^k}+\mathcal{P}_{\Lambda^k}^{\perp}\widetilde{\mathbf{z}}$ indicate the k-th reconstructed signal and the residual part, respectively.

Evidently, the noise projection term $(\widetilde{\mathbf{\Theta}}_{\Lambda^k}^{\mathrm{H}}\widetilde{\mathbf{\Theta}}_{\Lambda^k})^{-1}\widetilde{\mathbf{\Theta}}_{\Lambda^k}^{\mathrm{H}}\widetilde{\mathbf{z}}$ in (4.57) is merged to recovered signal and may significantly impair the performance of the aforementioned block greedy algorithm. In contrast, by the period estimation method in PBOMP, the potential subblock can be fast extracted from received signals without experiencing the process of choosing the support set one by one. Hence, the transmission error can be suppressed and modified with period estimation in PBOMP.

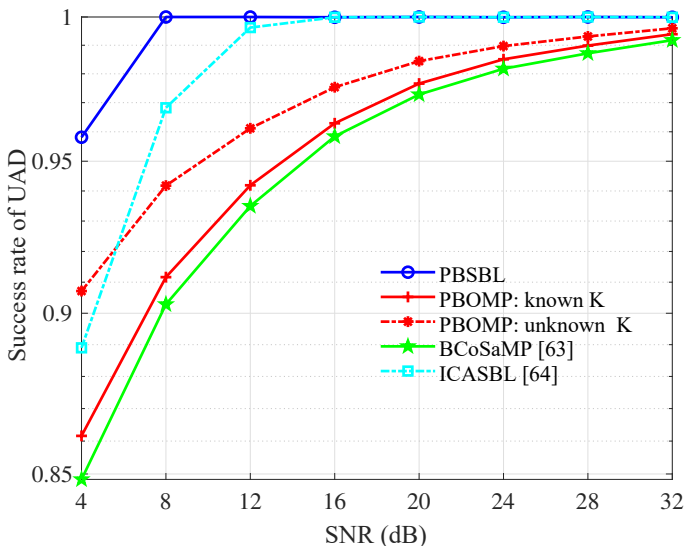

Figure 4.2: Success rate of UAD for the PBOMP, the PBSBL and comparison schemes.

The BER performance for different approaches is given in Figure 4.3. The PBSBL shows the best BER performance among different algorithms in the given SNR regime and finally reduces to 0 at SNR = 20 dB, which strongly confirms the effective suppression of

noise for the proposed PBSBL. Refer to (4.57), the false detection of active users leads to the significant transmission error in the procedure of signal reconstruction, and hence the proposed PBOMP without knowing the number of active users shows a lower BER performance than others. Nevertheless, the BER performance of PBOMP with knowing the number of active users is better than the BCoSaMP, which verifies the superiority of PBOMP in terms of signal reconstruction compared to the typical greedy algorithms (*e.g.*, BCoSaMP).

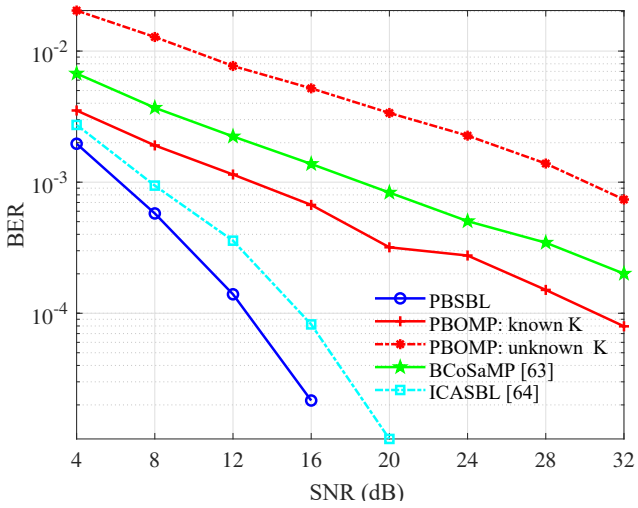

Figure 4.3: BER performance of the PBOMP, the PBSBL and comparison schemes.

Figure 4.4 depicts the NMSE performance comparison of the algorithms against SNR. Furthermore, the oracle LS method is shown as a benchmark to compare different algorithms. The performance of PBSBL with the prior information is superior to other algorithms over the whole SNR and the performance of PBSBL and PBOMP is better than ICASBL and BCoSaMP, respectively. On the other hand, the performance of PBOMP and PBSBL verify their superiority compared to their counterparts. This is because a significant number of subblock is effectively extracted from periodic block-sparsity structure, and such block sparsity feature can be fully exploited owing to period estimation. It also observed that the proposed PBSBL and Oracle LS method have close NMSE performance at a low SNR regime, which indicates the proposed PBSBL is robust to noise.

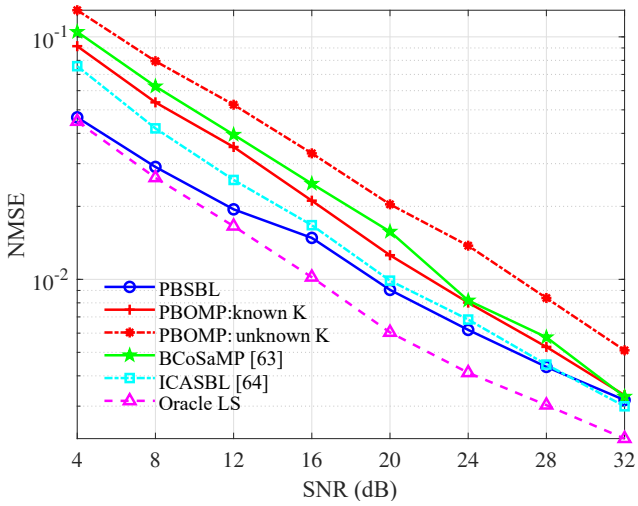

Figure 4.4: NMSE performance of the channel estimation for the PBOMP, the PBSBL and comparison schemes.

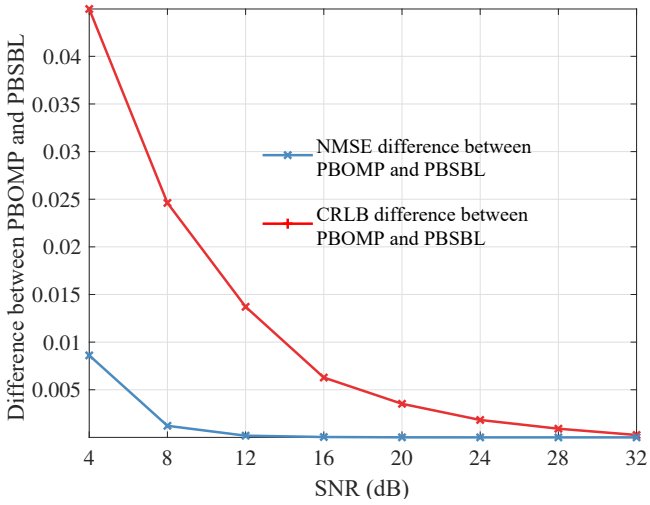

Figure 4.5: NMSE difference and the CRLB difference between the PBOMP and PBSBL.

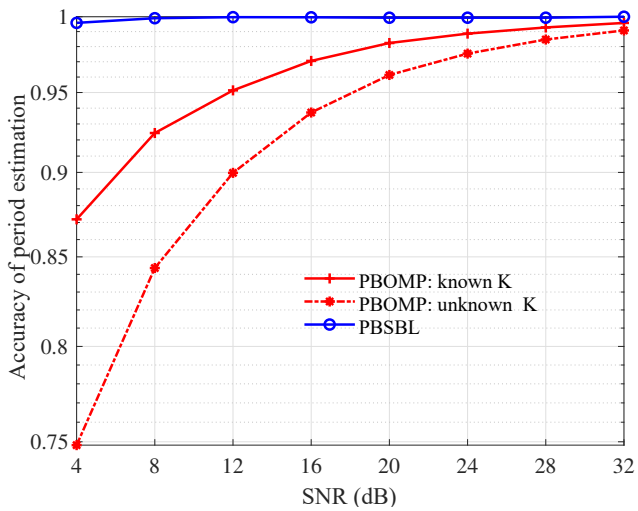

Figure 4.6: Performance of period estimation for the PBOMP and PBSBL.

The CRLB difference and NMSE difference of the channel estimation between the proposed PBOMP and PBSBL are presented in Figure 4.5. Note that in the proposed PBSBL, the NMSE of the channel estimation is related to the second term of the CRLB of the channel estimation, *i.e.*, $\beta_i \mathbf{D}_i$ in (4.37). Since β_i and $\mathbf{D}_i$ are iterative parameters to be estimated at each iteration of the PBSBL. The performance of the NMSE of the channel estimation for the PBSBL is affected by the performance of β_i and $\mathbf{D}_i$, which may lead to the gap between the NMSE difference and CRLB difference in practice. Nevertheless, the normalized CRLB difference and the normalized NMSE difference indicate that the channel estimation performance of the PBSBL is superior to the PBOMP in a given SNR region, which is in line with the analysis of Section 4.4. In particular, it is observed that the gap between the NMSE difference and the CRLB difference decreases with the increase of SNR. Hence, it is concluded that the PBSBL shows a close channel estimation performance to the PBOMP algorithm in a less noisy environment.

The proposed PBOMP and PBSBL for different SNR cases in terms of period estimation is firstly introduced in Figure 4.6. The simulation results show the proposed approaches achieve a favorable accuracy of period estimation, especially for moderate-to-high SNR. However, the performance of PBOMP without prior information deteriorates at low SNR environments. This is because the iteration condition in **Algorithm 2** relies on the accuracy of the estimated residual part, which is inevitably affected by noise. The accuracy of

period estimation for PBOMP with prior information K, in turn, runs at more than 98% at SNR = 30 dB. The PBSBL remains at a high level and keeps superior to the PBOMP among the whole SNR, which shows the PBSBL can effectively suppress the noise than the PBOMP.

Figure 4.7 shows the success rate of UAD with different sparsity degrees. On the whole, the performance of the success rate of UAD improves with the decrease of active number for the PBOMP and the PBSBL. In particular, the greater the number of active numbers, the greater the performance gaps between the proposed PBOMP with/without prior information. That is because the increase of the number of iteration in **Algorithm 2** induced by the increased active users may lead to more iterative errors in (4.14).

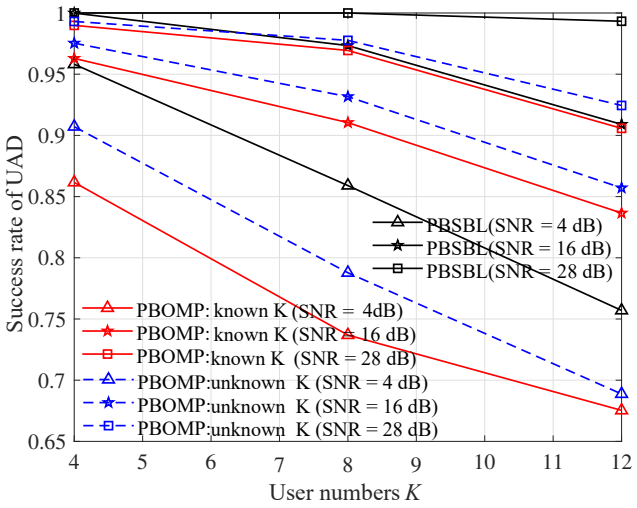

Figure 4.7: Impact of different numbers of active users (*e.g.*, 4, 8 and 12) on the UAD performance of the PBOMP and PBSBL with different SNRs.

Figure 4.8 shows the impact of sparsity degree on the accuracy of period estimation, where SNR = 4 dB, SNR = 16 dB and SNR = 28 dB are considered. When the active number is equal to 8, the PBOMP without prior information shows a sharp drop in terms of the performance of accuracy of period estimation. Furthermore, it is observed that the increase of the active number leads to a deteriorated performance of period estimation for PBOMP. However, this trend is less obvious for PBSBL. Hence, the PSBSBL can support

more users than the PBOMP.

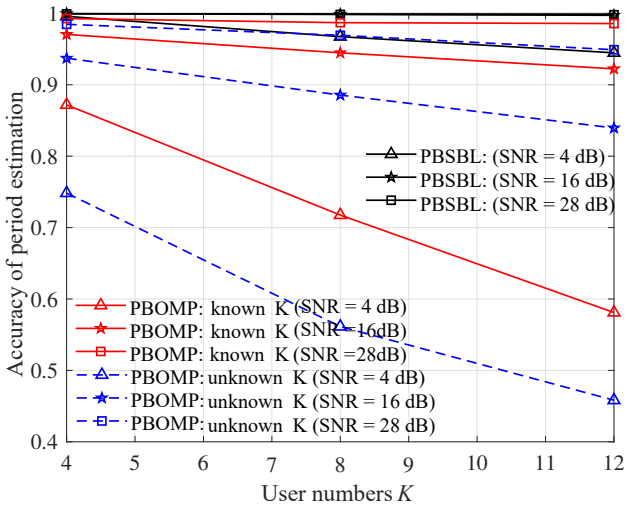

Figure 4.8: Impact of different numbers of active users (*e.g.*, 4, 8 and 12) on the performance of period estimation for the PBOMP and PBSBL with different system SNRs.

Figure 4.9 depicts the impact of different SNR scenarios (*e.g.*, 4 dB, 12 dB, and 20 dB) on the coefficient matrix for the proposed PBSBL. As an example, the block length and the number of active users are set to be 9 and 6, respectively. Furthermore, the upper triangle region of different blocks is presented in sequence, in which X-axis is the index of the elements in the upper triangular block, and Y-axis represents the index of the detection block. It can be observed that the correlation coefficient of noise blocks fluctuates greatly at a low SNR regime and the coefficients approach zero with the increase of SNR. In particular, the larger coefficients of noise block can hardly be recognized at high SNR. The coefficients of the signal block, by contrast, keep a high recognition at SNR regimes.

Figure 4.10 shows the impact of sparsity degree on the performance of the coefficient matrix at SNR = 12 dB, where signal block and adjacent noise block are randomly extracted. Apparently, the coefficients of the signal block are greater than that of the noise block for different active users. Since the performance of period estimation depends on the coefficients matrix, this result is consistent with Figure 4.8, where the performance of period estimation of PBSBL keeps stable for different active numbers. Hence, this property can be used to eliminate redundant information and capture periodic signals.

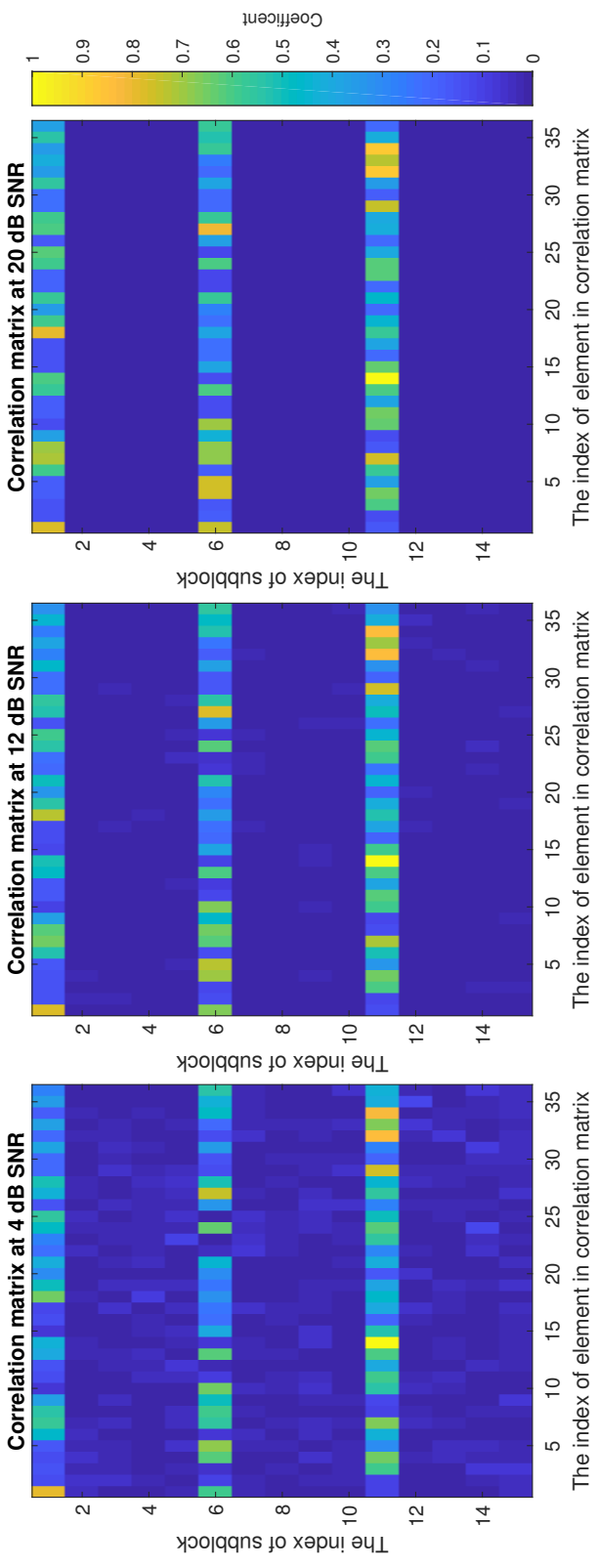

Figure 4.9: Impact of different system SNR (*e.g.*, 4 dB, 12 dB, and 20 dB) on the performance of coefficient matrix for the proposed PBSBL, where activation rate of users is set to be 15%.

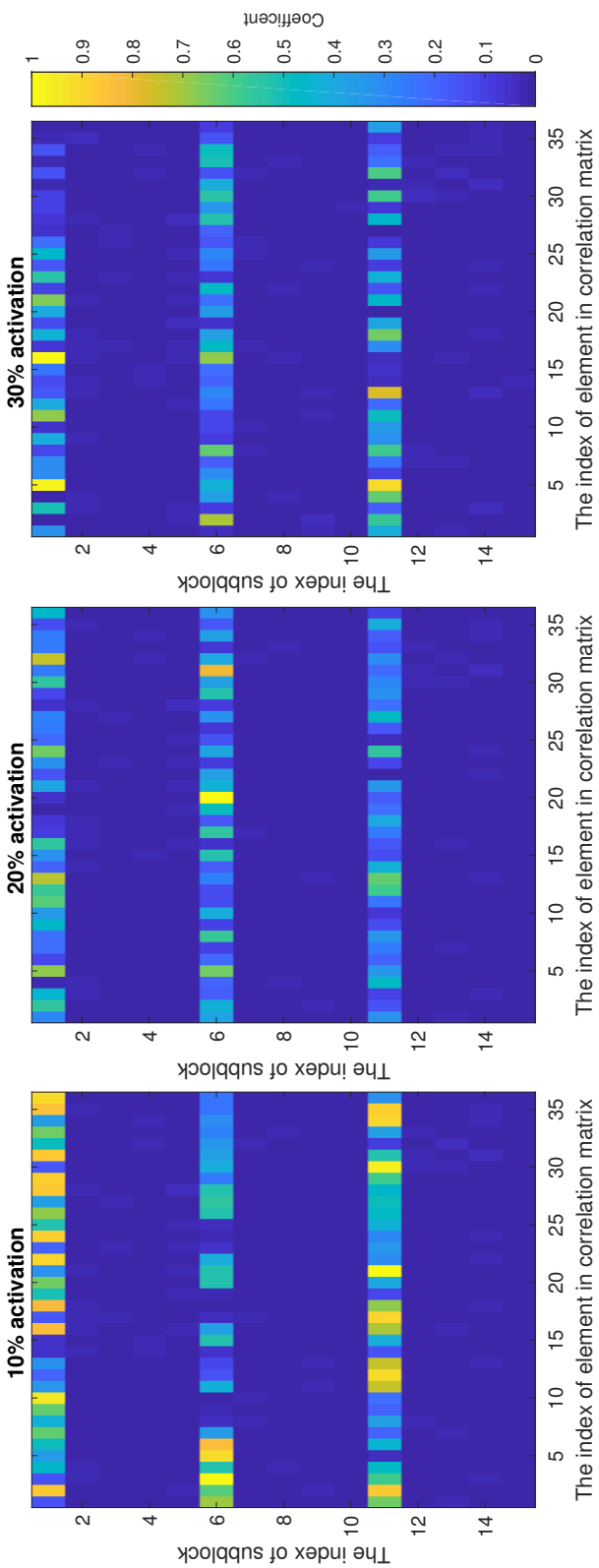

Figure 4.10: Impact of different activation rates of users (*e.g.*, 10%, 20% and 30%) on the performance of coefficient matrix for the proposed PBSBL, where the system SNR is set to be 12 dB.

4.7 Summary

In this chapter, a grant-free IoT system is considered, where only a fraction of users are simultaneously active and periodically transmit signals to an AP. The number of active users is unknown to AP. Building upon this scenario, a periodic block-sparsity structure has been presented, where two joint CS based channel estimation and signal detection approaches aided by period estimation, namely PBOMP and PBSBL, are proposed, respectively. By fully exploiting the non-continuous temporal correlation of the received data in the context of the periodic block-sparsity structure, the proposed approaches can effectively recover the transmitted signals from the superimposed signals without knowing the sparsity of active users at AP. The proposed PBOMP has low complexity, while the PBSBL is more robust against noise. To better understand the proposed approaches, their CRLBs are derived and compared. It is shown that the CRLB for channel estimation of PBSBL is lower than PBOMP when the noise power is greater than half of the power of the MA signature. In particular, the PBSBL enjoys a lower CRLB of channel estimation than PBOMP if BPSK or its phase-rotated version is adopted as a modulation scheme. The experimental results verified that the two proposed approaches are superior to the comparison algorithm in terms of the performance of period estimation and signal reconstruction. Furthermore, the iteration number of the PBOMP can be significantly reduced, and the abundant blocks can be fast removed by PBSBL based on period estimation.

The proposed PBOMP and PBSBL focused on periodic communication and assumed that there are only burst users. However, with the exponentially increasing number of devices, there are not only burst users, but also the users that already have live connection to the BS. The connected users may introduce interference to the detection of the burst users. Hence, in the following chapter, UAD and signal detection for grant-free communication are investigated in the presence of massive connected users.

Chapter 5

Burst User Detection for Grant-Free Communication in the Presence of Massive Connected Users

5.1 Introduction

Grant-free communication [17, 81, 94], which does not require a handshake process, is an effective approach to reducing the latency and signaling overhead imposed by the conventional request-grant access mechanism. As described in Chapter 2, one of the main challenges of grant-free communication lies in the massive users that already have live connection to BS may introduce interference to the detection of the burst users [95, 96].

In practice, the number of connected users is usually much larger than the number of to-be-connected burst users, imposing strong interference in UAD and signal detection. Hence, in this chapter, UAD and signal detection for grant-free communication are investigated in the presence of massive connected users and multiple antennas at the BS. The contributions in this chapter are as follows.

- This is the first reported work on interference suppression (IS) in grant-free communication. The number of effective interfering connected users is reduced to only one, by applying a preconditioning matrix to the received signals at BS. The remaining

interferer, referred to as the signature user, is selected to maximize of the burst users' SINR.

- An iterative UAD and SD scheme is applied after IS, where the prior information about the signature interfering user is exploited and the symbols of the burst users as well as the signature connected user are simultaneously detected by iterative exchanging of the soft information of user activity and symbols. The proposed scheme outperforms the existing algorithms in [71] and [80] in terms of the success rate of UAD and BE).

5.2 System Model

As shown in Figure 5.1, we consider an uplink grant-free communication, where the BS is equipped with R antennas for signal reception, and each user is equipped with a single antenna. Considering a massive-users scenario, an overloaded uplink transmission is assumed. Explicitly, signals generated by each user are transmitted on the same physical resources (*e.g.*, time- and frequency-domain), and the total number of users V is much larger than the number of antennas R of the BS. It is assumed that there are M users who have set up the connection with the BS, and let N denote the number of potential burst users, where we have $M + N = V$. Simultaneously, there are K_i ($K_i \ll N$) burst users trying to establish connection with the BS at the i-th slot, where $N = V - M$. Due to the sparsity transmission, the number of the burst users K_i is far less than the total number of users V [68, 96].

x

The received signal vector $\mathbf{y}$ at the BS can be expressed as

$$\mathbf{y} = \mathbf{H}_{\mathbb{S}_\mathrm{C}} \mathbf{\Xi}_{\mathbb{S}_\mathrm{C}} \mathbf{x}_{\mathbb{S}_\mathrm{C}} + \mathbf{H}_{\mathbb{S}_\mathrm{A}} \mathbf{\Xi}_{\mathbb{S}_\mathrm{A}} \mathbf{x}_{\mathbb{S}_\mathrm{A}} + \mathbf{z}, \tag{5.1}$$

where the under-scripts $\mathbb{S}_\mathrm{C}$ and $\mathbb{S}_\mathrm{A}$ denote the set of the connected users and potential burst users, respectively. $\mathbf{\Xi}_{\mathbb{S}_\mathrm{A}} = \mathrm{diag}(\sqrt{p_{\mathrm{A},1}}, ..., \sqrt{p_{\mathrm{A},N}})$ and $p_{\mathrm{A},n}$ is the transmission power of the n-th potential burst user. $\mathbf{\Xi}_{\mathbb{S}_\mathrm{C}} = \mathrm{diag}(\sqrt{p_{\mathrm{C},1}}, ..., \sqrt{p_{\mathrm{C},M}})$ and $p_{\mathrm{C},m}$ is the transmission power of the m-th connected user. $\mathbf{H}_{\mathbb{S}_\mathrm{C}} = [\mathbf{h}_{\mathrm{C},1}^\mathrm{T}, ..., \mathbf{h}_{\mathrm{C},M}^\mathrm{T}]^\mathrm{T} \in \mathcal{C}^{R \times M}$ and $\mathbf{H}_{\mathbb{S}_\mathrm{A}} = [\mathbf{h}_{\mathrm{A},1}^\mathrm{T}, ..., \mathbf{h}_{\mathrm{A},N}^\mathrm{T}]^\mathrm{T} \in \mathcal{C}^{R \times N}$. $\mathbf{h}_{\mathrm{C},m} \in \mathcal{C}^{R \times 1}$ denotes the wireless channel between the m-th connected user and the BS, while $\mathbf{h}_{\mathrm{A},n} \in \mathcal{C}^{R \times 1}$ denotes the wireless channel between the n-th burst user and the BS. $\mathbf{x}_{\mathbb{S}_\mathrm{C}} = [x_{\mathrm{C},1}, ..., x_{\mathrm{C},M}]^\mathrm{T} \in \mathcal{C}^{M \times 1}$ and

$\mathbf{x}_{\mathbb{S}_\mathrm{A}} = [x_{\mathrm{A},1}, ..., x_{\mathrm{A},n}]^\mathrm{T} \in \mathcal{C}^{N\times 1}$, where x_m and x_n are the data transmitted by the m-th connected user and the n-th potential active user, respectively. $\mathbf{z} \in \mathcal{C}^{R\times 1}$ is the noise vector and follows the complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \mathbf{1}_R)$ [48].

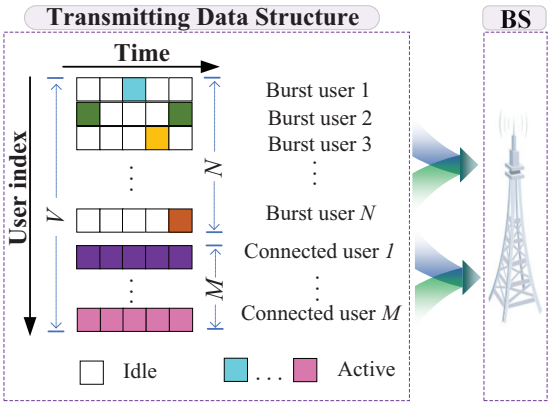

Figure 5.1: A typical uplink communication consists of M connected users and N potential active users, where the burst users become randomly active in different time slots, while the connected users keep active over the entire time slots.

In grant-free communication, their channel information can be readily estimated from the uplink reference signals. Furthermore, the channels are assumed under the block-fading, meaning that the channel matrix is invariant in a transmission block. In this context, the channel information of the connected users can be considered as the prior knowledge on the presence of massive users, which is used for the UAD and SD design in this thesis. Therefore, the channel information of different users is assumed to be readily known by the BS [47, 68, 71].

In this chapter, we first consider the scenarios that the users dissipate equivalent transmission power for sending signal. [47, 68, 71, 97]. Hence, the transmission power in the received signal vector is omitted for brevity, written as,

$$\mathbf{y} = \mathbf{H}_{\mathbb{S}_\mathrm{C}}\mathbf{x}_{\mathbb{S}_\mathrm{C}} + \mathbf{H}_{\mathbb{S}_\mathrm{A}}\mathbf{x}_{\mathbb{S}_\mathrm{A}} + \mathbf{z}. \tag{5.2}$$

5.3 Iterative UAD and Signal Detection with Interference Suppression

In this section, the interference suppression is given in Subsection 5.3.1. As shown in Figure 5.2, an iterative UAD and SD scheme is proposed in Subsection 5.3.2, where the prior information of massive connected users after the IS is delivered to the iterative process of exchanging soft information of the UAD and symbol information. Finally, the SINR based optimization of the IS is introduced in Subsection 5.3.3.

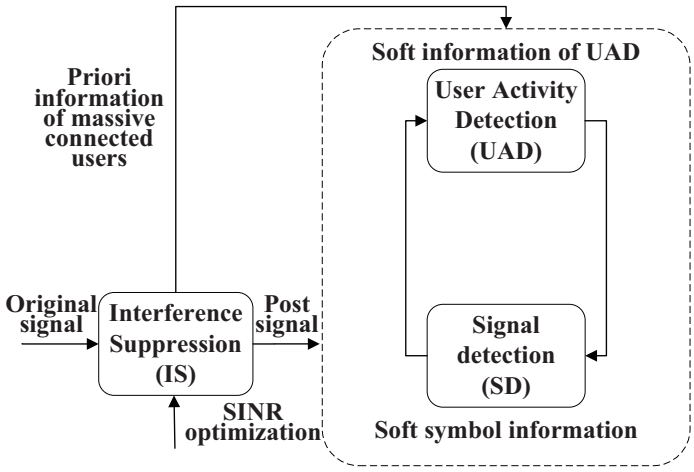

Figure 5.2: Proposed grant-free communication framework in the presence of massive connected users.

5.3.1 Interference Suppression

The received data at the BS comes from the users having connected to the BS as well as the burst users. As a result, when performing UAD for the burst user, it is imperative to reduce the interference induced by the connected users. To that end, we utilize the channel information of the connected users to obtain a preconditioning matrix to subtract the interference caused by the connected users. In detail, singular value decomposition is employed to extract the preconditioning matrix from the connected users [98], written as

$$\widetilde{\mathbf{H}}_{\mathrm{C}_{m^*}} = \left[\mathbf{U}^1_{m^*}, \mathbf{U}^2_{m^*}\right] \left[\mathbf{\Lambda}_{m^*}, \mathbf{0}\right] \left[\mathbf{V}_{m^*}\right], \tag{5.3}$$

where C_{m^*} denotes the m^*-th connected user, referred to as the signature user, and $\widetilde{\mathbf{H}}_{\mathrm{C}_{m^*}} = [\mathbf{h}_{\mathrm{C},1},...,\mathbf{h}_{\mathrm{C},m^*-1},\mathbf{h}_{\mathrm{C},m^*+1},...,\mathbf{h}_{\mathrm{C},M}]$. $\mathbf{U}^2_{m^*}$ can be extracted from left singular matrix of $\widetilde{\mathbf{H}}_{\mathrm{C}_{m^*}}$ and is spanned by $(R-M)$ vectors of null space of $\widetilde{\mathbf{H}}_{\mathrm{C}_{m^*}}\widetilde{\mathbf{H}}^{\mathrm{H}}_{\mathrm{C}_{m^*}}$. Hence, we have

$$\sum_{m\in\mathbb{S}_{\mathrm{C}},m\neq m^*} \mathbf{U}^{\mathrm{H}}_{m^*}\mathbf{h}_m x_m = \mathbf{0}, \tag{5.4}$$

where $\mathbf{U}_{m^*} = \mathbf{U}^2_{m^*}$. (5.4) shows that the signal generated by the m-th connected user can be eliminated by the matrix $\mathbf{U}_{m^*}$, where $m \in \mathbb{S}_{\mathrm{C}}$ and $m \neq m^*$. Hence $\mathbf{U}_{m^*}$, referred to as preconditioning matrix, is utilized to subtract interference from the connected users. Accordingly, the source of the interfering signal is reduced to only one. In other words, though there are M massive connected users, the interference from the M-1 connected users can be subtracted from the received signal at the BS. The post-processing signal is written as

$$\tilde{\mathbf{y}} = \mathbf{U}^{\mathrm{H}}_{m^*}\mathbf{y} = \mathbf{U}^{\mathrm{H}}_{m^*}\mathbf{H}_{\mathbb{S}_{\mathrm{A}}}\mathbf{x}_{\mathbb{S}_{\mathrm{A}}} + \mathbf{U}^{\mathrm{H}}_{m^*}\mathbf{h}_{m^*}x_{m^*} + \mathbf{U}^{\mathrm{H}}_{m^*}\mathbf{z}, \tag{5.5}$$

5.3.2 Iterative UAD and Signal Detection

After eliminating the interference from the connected users, the objective of grant-free communication is to recover the transmitted signals $\mathbf{x} = [\mathbf{x}_{\mathbb{S}_{\mathrm{C}}}\ \mathbf{x}_{\mathbb{S}_{\mathrm{A}}}] \in \mathcal{C}^{V\times 1}$ from the superimposed signals $\tilde{\mathbf{y}}$. Accordingly, the problem formulation is given as

$$\begin{aligned} &\text{P1: } \min_{\mathbf{x}} \left\|\tilde{\mathbf{y}} - \mathbf{U}^{\mathrm{H}}_{m^*}\mathbf{H}_{\mathbb{S}_{\mathrm{C}}}\mathbf{x}_{\mathbb{S}_{\mathrm{C}}} - \mathbf{U}^{\mathrm{H}}_{m^*}\mathbf{H}_{\mathbb{S}_{\mathrm{A}}}\mathbf{x}_{\mathbb{S}_{\mathrm{A}}}\right\|_2 \\ &\text{s.t. } (\mathrm{C1}) : \|\mathbf{x}_{\mathbb{S}_{\mathrm{A}}}\|_0 \ll \|\mathbf{x}_{\mathbb{S}_{\mathrm{C}}}\|_0 < V; \\ &\qquad (\mathrm{C2}) : R < V. \end{aligned} \tag{5.6}$$

where (C1) denotes that the number of the burst users K_i and the number of connected users M are less than the total number of users V. (C2) denotes an overloaded uplink transmission, where the number of the BS antennas R is less than the number of users V. At this point, it is ready to solve Problem P1 by examing the indexes of the burst users, which is equivalent to finding the burst user that corresponds to the maximum value of a *posteriori* activity probability.

. Lemma 1 *(A posteriori log-likelihood ratio (LLR) of the user activity): Assuming that alphabet of modulation symbol is ς, the posteriori LLR of the user activity for the n-th*

potential active user is defined as

$$L_n\left(\tilde{\mathbf{y}}\right) = \ln \frac{P(x_n \in \varsigma, x_{m^*} \in \varsigma|\tilde{\mathbf{y}})}{P(x_n = 0, x_{m^*} \in \varsigma|\widetilde{\mathbf{y}})} = L_{E,n}(\tilde{\mathbf{y}}) + L_{A,n}, \tag{5.7}$$

where

$$L_{E,n}(\tilde{\mathbf{y}}) = \ln \sum_{\varsigma_j} exp\ (L_{E,n,j}), \tag{5.8}$$

$$L_{A,n} = \ln \sum_{\varsigma_j} exp\ (L_{A,n,j}). \tag{5.9}$$

$L_{E,n,j}$ indicates the LLR of the j-th transmit symbol ς_j for the n-th user. $L_{A,n,j}$ denotes a prior LLR of the j-th transmit symbol ς_j for the n-th user. The derivations of $L_{E,n,j}$ and $L_{A,n,j}$ are detailed in Appendix A.1.

To find the burst users among the potential active users, we firstly define the detected index set of the burst users at the l-th iteration as $\mathbb{S}^{(l)}$ and define $\bar{\mathbb{S}}^{(l)}$ as the complementary set of $\mathbb{S}^{(l)}$. Then (5.7) can be employed to find the largest *posteriori* user activity probability, where the user corresponding to the maximum probability is the selected burst user, written as

$$n^* \approx \underset{n \in \bar{\mathbb{S}}^{(l-1)}}{\arg\max} (\max_j L_{E_1,n,j}(\tilde{\mathbf{y}}) + L_{A,n}), \tag{5.10}$$

where the proof of (5.10) is shown in Appendix A.2. In particular, $L_{E_1,n,j}(\tilde{\mathbf{y}})$ denotes the soft information of the previously selected users which will be delivered to SD. By doing so, the soft symbol information, as detailed later, can be updated and employed to perform SD. The interference-plus-noise vector can be regarded as a Gaussian distributed variable, and we have $\mathrm{E}_{\mathrm{x}_{\mathbb{S}_{\mathrm{A}}^{(l-1)},x_m}}\left[\exp\left(\cdot\right)\right] \approx \exp(\mathrm{E}_{\mathrm{x}_{\mathbb{S}_{\mathrm{A}}^{(l-1)},x_m}}\left[\cdot\right])$ [68]. Hence, $L_{E_1,n,j}(\tilde{\mathbf{y}})$ can be obtained by (5.11).

$$\begin{aligned} L_{E_1,n,j}(\tilde{\mathbf{y}}) &= \ln \frac{\mathrm{E}_{\mathrm{x}_{\mathbb{S}^{(l-1)},x_m}}\left[\exp\left(-\left\|\tilde{\mathbf{y}} - \mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{H}_{\mathbb{S}^{(l-1)}}\mathbf{x}_{\mathbb{S}^{(l-1)}} - \varsigma_k\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_n - x_{m^*}\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_{m^*}\right\|_{\Lambda_n^{(l)-1}}^2\right)\right]}{\mathrm{E}_{\mathrm{x}_{\mathbb{S}^{(l-1)},x_{m^*}}}\left[\exp\left(-\left\|\tilde{\mathbf{y}} - \mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{H}_{\mathbb{S}^{(l-1)}}\mathbf{x}_{\mathbb{S}^{(l-1)}} - x_{m^*}\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_{m^*}\right\|_{\Lambda_n^{(l)-1}}^2\right)\right]} \\ &= \Re\left\{\varsigma_j\left[\left(\tilde{\mathbf{r}} - \varsigma_j\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_n\right)^{\mathrm{H}}\Lambda_n^{(l)-1}\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_n + \left(\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_n\right)^{\mathrm{H}}\Lambda_n^{(l)-1}\tilde{\mathbf{r}}\right]\right\}. \end{aligned} \tag{5.11}$$

The matrix $\Lambda_n^{(l)}$ in (5.11) is given by

$$\Lambda_n^{(l)} = \mathrm{Cov}\left(\sum_{k\neq n, k\in\bar{\mathbb{S}}^{(l-1)}} \bar{x}_k \mathbf{U}_{m^*}^{\mathrm{H}} \mathbf{h}_k + \mathbf{U}_{m^*}^{\mathrm{H}} \mathbf{z}\right). \tag{5.12}$$

Furthermore, $\tilde{\mathbf{r}}^{(l)}$ in (5.11) indicates the residue of the received signal at the l-th iteration, expressed as

$$\tilde{\mathbf{r}}^{(l)} = \tilde{\mathbf{y}} - \sum_{k\in\mathbb{S}^{(l-1)}} \bar{x}_k \mathbf{U}_{m^*}^{\mathrm{H}} \mathbf{h}_k - \bar{x}_{m^*} \mathbf{U}_{m^*}^{\mathrm{H}} \mathbf{h}_{m^*}, \tag{5.13}$$

where $\bar{x}_k$ and $\bar{x}_{m^*}$ are the expectation of the data generated by the k-th potential active user and signature user, respectively. $\bar{x}_k$ is given by

$$\bar{x}_k = \frac{\exp\left(L_{A,k} - \ln|\varsigma|\right) \sum\limits_{\varsigma_j\in\varsigma} \exp(L_{E_2,k,j})\varsigma_j}{1 + \exp\left(L_{A,k} - \ln|\varsigma|\right) \sum\limits_{\varsigma_j\in\zeta} \exp(L_{E_2,k,j})}, \tag{5.14}$$

where the detailed derivation of $\bar{x}_k$ is given in Appendix A.3. In (5.14), $L_{E_2,k,j}$ denotes the soft symbol information of the detected burst users and is delivered from the process of SD. Based on the fact of $P_{m^*} = 1$ and (5.14), the expectation of the data of the signature user $\bar{x}_{m^*}$ in (5.13) is given by

$$\bar{x}_{m^*} = \frac{\sum\limits_{\varsigma_j\in\varsigma} \exp\left(L_{E_2,m^*,j}\right)\varsigma_j}{\sum\limits_{\varsigma_j\in\varsigma} \exp\left(L_{E_2,m^*,j}\right)}. \tag{5.15}$$

Since the maximum value of $L_{E_1,n,j}$ corresponds to the most probable symbol ς_j, it can be used to perform SD. In particular, after the $L_{E_1,n,j}$ is updated, a newly selected burst user is merged to the active set. Here we denote the updated $L_{E_1,n,j}$ as $L_{E_2,k,j}$ for clarity. Following the guideline, the $L_{E_2,k,j}$ is given by (5.16) as shown on the top of next page, where $\Gamma^{(l)} = \mathbb{S}^{(l)} - \{n\}$ and the matrix $\Upsilon_n^{(l)}$ at the l-th iteration can be given by

$$\Upsilon_n^{(l)} = Cov\left(\sum_{k\in\mathbb{S}^{(l)}} \bar{x}_k \mathbf{U}_{m^*}^{\mathrm{H}} \mathbf{h}_k + \mathbf{U}_{m^*}^{\mathrm{H}} \mathbf{z}\right). \tag{5.17}$$

$$L_{E2,n,j}(\tilde{\mathbf{y}}) = \ln \frac{\mathrm{E}_{\mathrm{x}_{\Gamma^{(l)}},x_{m^*}}\left[\exp\left(-\left\|\tilde{\mathbf{y}}-\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{H}_{\Gamma^{(l)}}\mathbf{x}_{\Gamma^{(l)}}-\varsigma_k\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_n-x_{m^*}\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_{m^*}\right\|^2_{\Upsilon_n^{(l)-1}}\right)\right]}{\mathrm{E}_{\mathrm{x}_{\Gamma^{(l)}},x_{m^*}}\left[\exp\left(-\left\|\tilde{\mathbf{y}}-\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{H}_{\Gamma^{(l)}}\mathbf{x}_{\Gamma^{(l)}}-x_{m^*}\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_{m^*}\right\|^2_{\Upsilon_n^{(l)-1}}\right)\right]}$$
$$= \Re\left\{\varsigma_j\left[\left(\tilde{\mathbf{r}}-\varsigma_j\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_n\right)^{\mathrm{H}}\Upsilon_n^{(l)-1}\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_n+\left(\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{h}_n\right)^{\mathrm{H}}\Upsilon_n^{(l)-1}\tilde{\mathbf{r}}\right]\right\}. \tag{5.16}$$

Similar to (5.14), the updated $\bar{x}_k$ in (5.17) is calculated by

$$\bar{x}_k = \frac{\exp\left(L_{A,n}-\ln|\varsigma|\right)\sum\limits_{\varsigma_j\in\varsigma}\exp(L_{E,k,j})\varsigma_j}{1+\exp\left(L_{A,n}-\ln|\varsigma|\right)\sum\limits_{\varsigma_j\in\varsigma}\exp(L_{E,k,j})}, \tag{5.18}$$

where

$$L_{E,k,j} = \begin{cases} L_{E_1,k,j} & \text{if } k \overset{\text{(a)}}{=} n^*, \\ L_{E_2,k,j} & \text{otherwise.} \end{cases} \tag{5.19}$$

Once a burst user (*e.g.,* the n^*-th user) is captured, the soft information $L_{E_1,k,j}$ is merged into soft symbol information of SD $L_{E_2,k,j}$. In this context, condition (a) in (5.19) is performed to update soft symbol information of SD $L_{E_2,k,j}$ for the next UAD process.

Since the number of the burst users K_i is unknown at the BS, a termination condition should be imposed to avoid the reconstructed error caused by the idle users. Assuming the maximum number of iterations L, the power of the residual signal equals to the power of noise when the support set $\mathbb{S}^{(L)}$ is perfectly recovered. Accordingly, the termination condition is given by

$$\left\|\tilde{\mathbf{r}}^{(l-1)}\right\|_2^2 \le \left\|\tilde{\mathbf{r}}^{(l)}\right\|_2^2 \text{ or } \left\|\tilde{\mathbf{r}}^{(l)}\right\|_2^2 \le \left\|\mathbf{U}_{m^*}^{\mathrm{H}}\mathbf{z}\right\|_2^2. \tag{5.20}$$

After the active users and their transmitted signals are collected, the signal generated by the m-th connected user is detected by the successive interference cancellation and Least

Square method, which can be expressed as

$$x_m = \min_{x_m \in \varsigma} \left\| \mathbf{U}_m^{\mathrm{H}} \mathbf{y} - \mathbf{U}_m^{\mathrm{H}} \mathbf{H}_{\mathbb{S}^{(l)}} \mathbf{x}_{\mathbb{S}^{(l)}} + \mathbf{U}_m^{\mathrm{H}} \mathbf{h}_{\mathrm{m}} x_{\mathrm{m}} + \mathbf{U}_m^{\mathrm{H}} \mathbf{z} \right\|_2^2 . \tag{5.21}$$

Now we are able to devise the whole algorithm. The proposed approach is employed by exchanging the soft information between UAD and SD through an iterative process. Specifically, by substituting (5.11) into (5.8), the potential active users at the l-th iteration can be collected into a set $\mathbb{S}^{(l)}$. Subsequently, the set $\mathbb{S}^{(l)}$ corresponding to soft information $L_{E_1,n,j}$ is delivered to the soft symbol information, *i.e.*, $L_{E_2,n,j}$ in the procedure of SD. By doing so, the detected data from the collected active users in line with the set $\mathbb{S}^{(l)}$ can be iteratively updated. After the data information is updated, the soft information $L_{E_2,n}$ is delivered to $L_{E_1,n,j}$ to find the active users. The proposed iterative UAD and SD approach is summarized in **Algorithm 4**.

Algorithm 4 Iterative UAD and signal detection Algorithm

1: Select a signature user and design preconditioning matrix by (5.3);
2: **for** $l = 1$ to L **do**
3: **User Activity Detection**:
4: Update soft information $L_{E_1,n,j}$ by (5.11);
5: Find the index of active user by (5.10);
6: **Signal Detection**:
7: Update soft symbol information $L_{E_2,n,j}$ by (5.16);
8: Quit iteration if (5.20) is satisfied
9: **end for**
10: Signal detection for connected users by (5.3) and (5.21).

5.3.3 Signature User Selection

In Subsection 5.3.2, we have proposed an iterative UAD and SD algorithm, assuming that users employ identical transmission power for sending signal. In this subsection, we further consider a more practical scenario, where users may dissipate different levels of transmission power. We aim to find a signature user with different transmission power. The selected signature user corresponds to a preconditioning matrix, which minimizes the interference caused by the connected users. To this end, the SINR of the burst users directly determines the option of the signature user. In other words, if a connected user can maximize the SINR of active users by the preconditioning matrix $\mathbf{U}_m$, it can be selected as the signature

user. Hence, referring to (5.3), the index of the signature user is acquired by

$$m^* = \arg\max_{m \in \mathbb{S}_{\mathrm{C}}} \frac{\mathrm{E}\left\|\mathbf{H}_{\mathbb{S}_{\mathrm{A}}} \mathbf{\Xi}_{\mathbb{S}_{\mathrm{A}}} \mathbf{x}_{\mathbb{S}_{\mathrm{A}}}\right\|_2^2}{\mathrm{E}\left\|\mathbf{H}_{\mathbb{S}_{\mathrm{A}}} \mathbf{\Xi}_{\mathbb{S}_{\mathrm{A}}} \mathbf{x}_{\mathbb{S}_{\mathrm{A}}}\right\|_2^2 + \mathrm{E}\left\|\sqrt{p_{\mathrm{C},m}} \mathbf{H}_m x_m\right\|_2^2 + \mathrm{E}\left\|\mathbf{z}\right\|_2^2}. \tag{5.22}$$

(5.22) is based on the fact that the preconditioning matrix $\mathbf{U}_m$ is independent with burst users, the m-th connected user and noise.

5.4 Simulation Results

In this section, the UAD and SD performance of the proposed iterative algorithm for UAD and SD is demonstrated. Two metrics are applied to evaluate the accuracy of the signal reconstruction procedure, consisting of the success rate of UAD and BER. The success rate of UAD denotes the ratio between the numbers of the detected burst users and the total number of active users. The total number of users V, the number of antennas at the BS R, the number of the connected users M are set as 96, 64 and 32, respectively. The number of the burst users K_i at the i-th slot accounts for 10% of the total number of potential burst users in Figure 5.3, 5.4 and 5.6, while the number of the burst users is increased in Figure 5.5 to evaluate the UAD performance in terms of the different number of burst users. Without loss of generality, transmission power is normalized to 1. The classical CS algorithm, *i.e.*, OMP [80], and EMP algorithms [71] are selected as benchmarks.

Performance of the success rate of UAD by the different algorithms is shown in Figure 5.3. It is shown that the performance of the benchmarks without preconditioning matrix is lower than that with preconditioning matrix. However, since the OMP and EMP algorithms rely on the strictly restricted isometry property (RIP) [99], the benchmarks with preconditioning matrix is still inferior to the proposed algorithm. Explicitly, the RIP measures the relationship between the sparse degree K_i, *i.e.*, the number of burst users, and the convergence of CS algorithms. Hence, the CS algorithm with a lower sparse degree could achieve better performance in terms of the reconstructed signal. By contrast, it is observed that the proposed approach can reach the perfect UAD performance with 16 dB SNR, meaning that the proposed approach provides a relaxed condition of RIP.

Figure 5.4 shows BER performance of different algorithms. The proposed approach significantly outperforms the benchmarks at all SNR regions, confirming its strong capability on interference suppression.

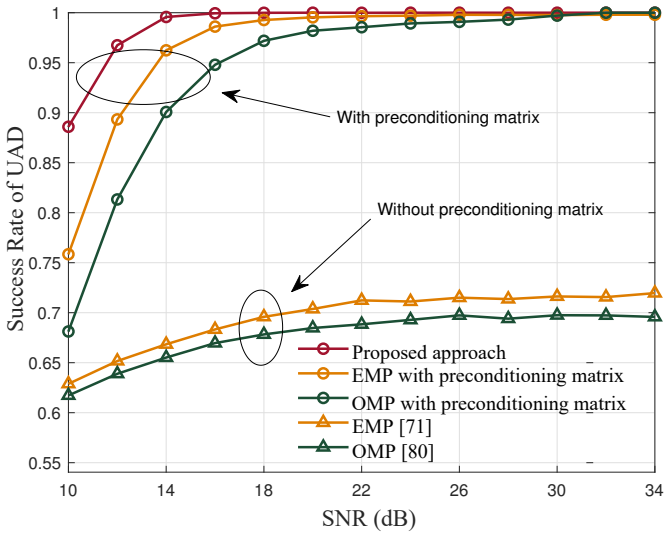

Figure 5.3: Impact of SNR value on success rate of UAD.

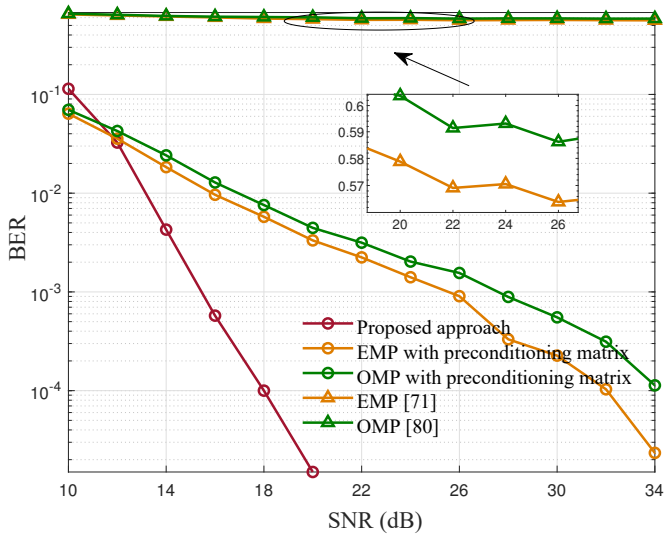

Figure 5.4: Impact of SNR value on BER performance.

To furthermore verify the robustness against the varying number of the burst users, Figure 5.5 is given to present UAD performance against different number of the burst users K_i, ranging from 1 to 6. It is observed that with more burst users at each slot, the UAD performance is slightly degraded, due to the increased level of uplink multiuser interference. However, in moderate/high SNR regions, the proposed approach can still obtain high UAD performance. Hence, it can be concluded that the proposed approach is applicable for arbitrary number of burst users K_i.

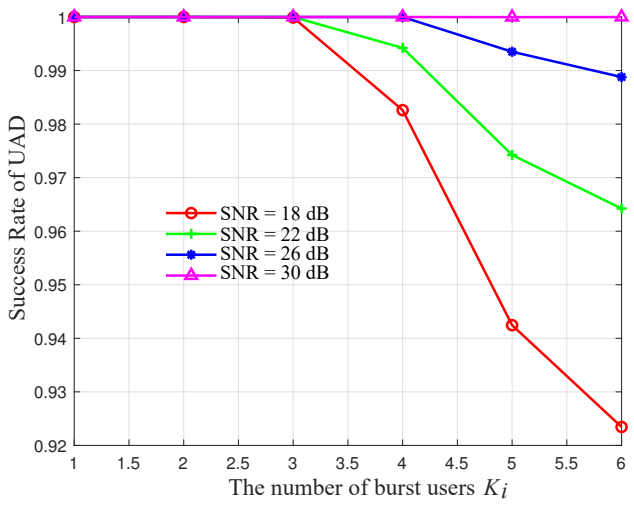

Figure 5.5: UAD performance against different number of burst users K_i.

Impact of selecting signature user on the accuracy of UAD is presented in Figure 5.6. As can be observed, by selecting the signature users that results in the minimum interference to the burst users, the UAD performance can be significantly enhanced. In contrast, random selection may return a signature user that imposes high residual interference to the burst users, thus degrading the UAD performance.

To compare the complexity of the proposed algorithm and the benchmarks, the elapsed time required by the different algorithms is recorded based on MATLAB R2020b and i7-8700K CPU platform. Since the preconditioning matrix can be calculated offline for a slow fading channel, the complexity of the preconditioning matrix is dismissed in practice. Hence, the main complexity of the proposed algorithm is composed of updating soft information $L_{E_1,n,j}$ and $L_{E_2,n,j}$. The simulation results show that the average elapsed times required by the proposed algorithm, OMP and EMP are 9.9 ms, 0.69 ms and 2.5

ms, respectively. Accordingly, the classical OMP algorithm enjoys the lowest complexity among the different algorithms, while the complexity of the proposed is slightly larger than the EMP. Furthermore, considering K users, Table 5.1 provides the complexities of the proposed detector and its counterparts.

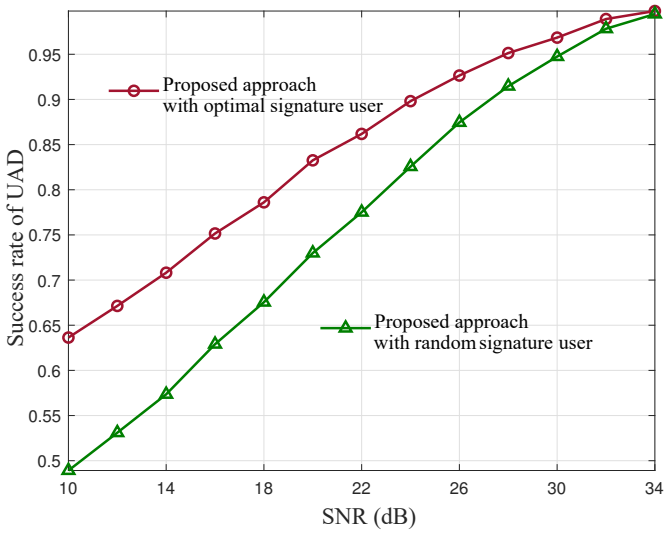

Figure 5.6: Impact of SNR on the success rate of UAD, where the users dissipate different level of transmission power.

Table 5.1: Complexity analysis of different algorithms.

Method	**Complexity**
OMP [80]	$O(RK^3)$
EMP [71]	$O(RK^3)$
Proposed Algorithm	$O(2K^3(R-M)^3)$

5.5 Summary

In this chapter, an iterative UAD and SD scheme in massive-user grant-free communication have been proposed. The prior information about the interfering users is exploited and the data from the connected and the burst users is simultaneously detected by iteratively exchanging of soft information regarding the user activity and symbols. Accordingly, the proposed approach is capable of providing significant performance gains over the existing algorithms in terms of the success of UAD and BER.

In Chapter 3, 4 and 5, CS based UAD schemes were proposed for the cases of aperiodic and periodic traffic, respectively, by employing the temporal correlation of the received signals. However, the aforementioned works assumed that the pilot matrix keeps unchanged over the entire transmission period, resulting in serious security problems when an eavesdropper has the knowledge of the pilot information. To this end, the following chapter focus on the secure channel estimation for grant-free communication.

Chapter 6

Channel Estimation for Secure Grant-Free Communication

6.1 Introduction

As described in Chapter 2, Chapter 3 and Chapter 4, the existing CS based channel estimation techniques depend on the specific pilot matrix. In other words, the constant pilot matrix may introduce serious security problems when an eavesdropper has the knowledge of the pilot information. Hence, how to design secure CS based channel estimation is one of the challenges of user detection for grant-free communication. Motivated by the aforementioned open challenges, in this chapter, an environment-aware (EA) pilot-based channel estimation is designed to ensure secure grant-free communication. The contributions of this chapter are given as follows.

- The standard CS forms of the channel estimation for the proposed approach are derived based on the time domain and virtual angular domain, respectively. Furthermore, the gain pattern of the uniform linear array (ULA) antenna with fixed direction angle of arrival (AoA) is given to show sparse signals in the virtual angular domain.

- Based on different domains, an EA pilot scheme is designed to meet the requirement of secure grant-free communication, where the channel information of different users can hardly be acquired by the eavesdropper. Furthermore, the proof of the validity of the pilot design, *i.e.*, measurement matrix, is given. The performance of the channel

estimation of the proposed scheme is presented based on the different sparse basis in LTE-A with the frequency selective channel.

6.2 System Model

A typical grant-free communication with a BS and N potential active users is considered, where the Time Division Duplex (TDD) communication protocol is employed to achieve the high frequency spectrum efficiency. Furthermore, NOMA is employed as an advanced radio access technology to fulfill the requirement of massive user connectivity. Without loss of generality, each user and BS is equipped with a single antenna and R antennas, respectively. As shown in Figure 6.1, there are N burst users who simultaneously wake up and transmit signals to the BS. An eavesdropper with the same number of antennas as the BS keeps monitoring the potential active users and tries to decrypt messages generated by active users.

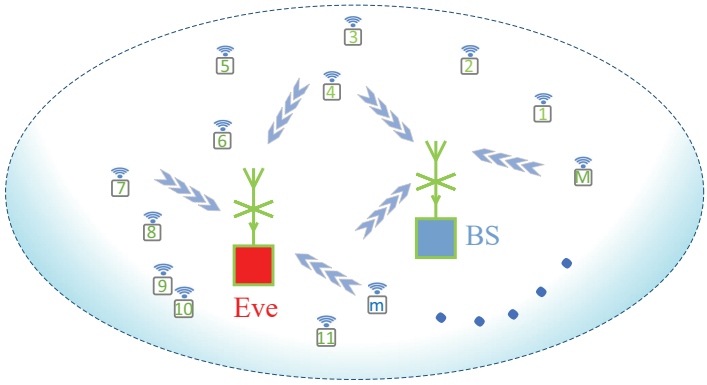

Figure 6.1: A typical attack from the eavesdropper for grant-free communication.

The channel impulse for the n-th user in the time domain can be modeled as

$$\mathbf{H}_n(\tau) = \sum_{l=0}^{L_n-1} \mathbf{H}_{l,n} e^{j2\pi f_d t} \delta(\tau - \tau_{l,n}), \tag{6.1}$$

where L_n, $\tau_{l,n}$ and f_d denotes the number of multipath, the l-th path delay and the

maximum Doppler shift, respectively. Furthermore, the amplitude and phase of the n-th user are given by

$$\mathbf{H}_{l,n} = h_{l,n}\boldsymbol{\psi}_{\mathrm{BS}(\theta_{l,n})}\boldsymbol{\psi}_{\mathrm{UE}(\phi_{l,n})}, \tag{6.2}$$

where $h_{l,n}$ denotes l-th path of complex gain. The azimuth AoA $\theta_{l,n}$ and the azimuth angle of departure $\phi_{l,n}$ follow the uniform distribution. Furthermore, steering vectors at the BS and user can be respectively expressed by

$$\begin{aligned}\boldsymbol{\psi}_{\mathrm{BS}}(\theta_{l,n}) &= [e^{\frac{j2\pi(0)dsin(\theta_{l,n})}{\lambda}}, \ldots, e^{\frac{j2\pi(R-1)dsin(\theta_{l,n})}{\lambda}}]^{\mathrm{T}}, \\ \boldsymbol{\psi}_{\mathrm{UE}}(\phi_{l,n}) &= [e^{\frac{j2\pi(0)dsin(\phi_{l,n})}{\lambda}}, \ldots, e^{\frac{j2\pi(U-1)dsin(\phi_{l,n})}{\lambda}}]^{\mathrm{T}},\end{aligned} \tag{6.3}$$

where U denotes the number of user's antenna. Accordingly and based on the frequency domain, (6.2) can be expressed by

$$\mathbf{H}^{f}_{p,n} = \sum_{l=0}^{L_n-1} \mathbf{H}_{l,n} e^{\frac{-j2\pi f_s \tau_{l,n}(1+f_d)p}{P}}, \tag{6.4}$$

where f_s and p denote sampling rate and the p-th of P pilots, respectively.

Since each user is equipped with only one antenna, *i.e.*, $U = 1$. Hence, steering vectors at the user side are formed into a unit matrix. In this chapter, the low-velocity service is considered, and hence, the small-scale follows a slowly time-varying channel.

6.3 EA-based Channel Estimation

In this section, CS based grant-free channel estimation in the time domain is firstly given in Subsection 6.3.1, followed by CS based grant-free channel estimation in the virtual angular domain in Subsection 6.3.2. Finally, a secure uplink grant-free transmission scheme and a novel pilot design is proposed to enhance the wireless security, which are shown in Subsection 6.3.3.

6.3.1 CS based Channel Estimation in Time Domain

The time channel impulse of the k-th orthogonal frequency division multiplexing (OFDM) symbol between the n-th user and the r-th antenna at the BS can be expressed as $\boldsymbol{h}_{r,n} = [h_{r,n}(0), h_{r,n}(1), \cdots, h_{r,n}(L_n-1)]^{\mathrm{T}}$, where L_n denotes maximum channel delay of the $n-$th

user. At the BS, the received pilot sequence of the r-th antenna can be written as

$$\mathbf{Y}_{k,r} = \sum_{n=0}^{N-1} \mathrm{diag}(\hat{\boldsymbol{\alpha}}_{k,n})\mathbf{F}_L \mathbf{h}_{r,n} + \mathbf{Z}, \tag{6.5}$$

where $\hat{\boldsymbol{\alpha}}_{k,n}$, $\mathbf{F}_L$ and $\mathbf{Z}$ denote pilot vector, the first L columns of the DFT matrix and AWGN, respectively. Furthermore, $L = \max(L_n), n \in (1, N)$.

6.3.2 CS based Channel Estimation in Virtual Angular Domain

Authors in [31] and [100] introduced the concept of the virtual angular domain, where the energy of the received signal is limited in a restricted region. Accordingly, the gain pattern of a uniform linear array antenna with fixed direction AoA is simulated as shown in Figure 6.2, where the AOA is set to 45 degrees with 16, 32, 64 and 128 antennas at the BS, respectively. It can be observed that the resolution of AoA in the virtual angular domain increases with the number of antennas. Furthermore, there are limited scatters around the BS, where the signal is sparse in the virtual angular domain with finite AoA. Given this guideline, the channel estimation of sparse users is detailed later in this subsection.

The classical communication theory depicts the explicit channel characteristics on different use cases, where coherence time quantifies the similarity of the channel impulse at different times [101]. Hence, the coherence time T_c in Rayleigh fading channel is expressed by

$$T_c = \sqrt{\frac{9}{16\pi f_d^2}}, \tag{6.6}$$

where simultaneously estimate consecutive symbols can be performed if transmission time is far less than the coherence time. Applying the multi-path slowly varying channels, the received pilot sequence of the p-th pilot of R antennas for K continuous transmission symbols at the BS can be written as

$$\mathbf{Y}_p = \sum_{n=0}^{N-1} \bar{\boldsymbol{\alpha}}_{n,p} \mathbf{h}_{n,p}^{\mathrm{T}} + \mathbf{Z}, \tag{6.7}$$

where $\mathbf{Y}_p = [\mathbf{y}_{0,p}, \ldots, \mathbf{y}_{K-1,p}]^{\mathrm{T}} \in \mathbb{C}^{K \times R}$, $\mathbf{y}_{k,p} = [y_{k,0,p}, \ldots, y_{k,R-1,p}]^{\mathrm{T}} \in \mathbb{C}^{R \times 1}$, $\bar{\boldsymbol{\alpha}}_{n,p} = [\bar{\alpha}_{0,n,p}, \ldots, \bar{\alpha}_{K-1,n,p}]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}$ and $\mathbf{h}_{n,p} = [h_{0,n,p}, \ldots, h_{R-1,n,p}]^{\mathrm{T}} \in \mathbb{C}^{R \times 1}$.

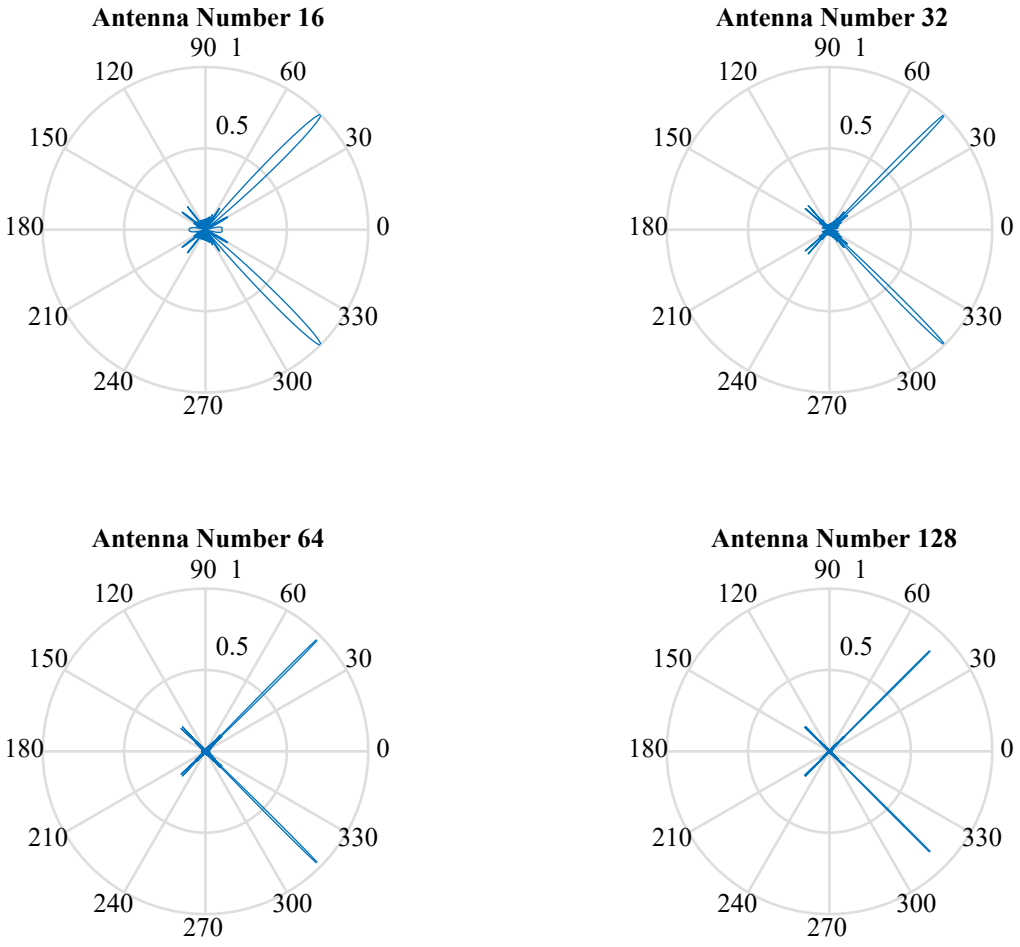

Figure 6.2: Fixed AOA and gain pattern of the ULA equipped with different number of antennas.

The channel matrix can be expressed as a sparse form, such as

$$\bar{\mathbf{h}}_{n,p} = \mathbf{A}_{\mathrm{BS}} \mathbf{h}_{n,p} \mathbf{A}_{\mathrm{UE}}, \tag{6.8}$$

where $\mathbf{A}_{\mathrm{BS}} = [\psi_{\mathrm{BS}}(0), \ldots, \psi_{\mathrm{BS}}(R-1)]$ and $\mathbf{A}_{\mathrm{UE}} = 1$. Hence, it can be concluded that

$$\begin{aligned} \mathbf{vec}(\mathbf{Y}_p) &= \sum_{n=0}^{M-1} \mathbf{A}_{\mathrm{BS}}^{\mathrm{T}} \otimes \bar{\boldsymbol{\alpha}}_{n,p} \bar{\mathbf{h}}_{n,p} + \mathrm{vec}(\mathbf{Z}) \\ &= \boldsymbol{\Phi}_p \mathbf{I} \bar{\mathbf{H}}_p + \mathrm{vec}(\mathbf{Z}) \\ &= \boldsymbol{\Theta}_p \bar{\mathbf{H}}_p + \mathrm{vec}(\mathbf{Z}), \end{aligned} \tag{6.9}$$

where $\boldsymbol{\Phi}_p = [\mathbf{A}_{\mathrm{BS}}^{\mathrm{T}} \otimes \bar{\boldsymbol{\alpha}}_{0,p}, \ldots, \mathbf{A}_{\mathrm{BS}}^{\mathrm{T}} \otimes \bar{\boldsymbol{\alpha}}_{N-1,p}]$. $\mathbf{I}$ denotes the unity matrix with a size of $NK \times RN$. Furthermore, $\bar{\mathbf{H}}_p = [\bar{\mathbf{h}}_{0,p}^{\mathrm{T}}, \ldots, \bar{\mathbf{h}}_{N-1,p}^{\mathrm{T}}]^{\mathrm{T}}$. Based on (6.8) and (6.9), channel information of sparse users can be obtained [72].

6.3.3 Pilot Design

Based on Subsection 6.3.1 and 6.3.2, it is clear that the eavesdropper needs to know all users' pilots to acquire channel information of active users. In other words, grant-free communications can not be conducted under a blanket of secrecy once the pilot book is disclosed. Hence, the unchanged users' pilots may lead to severe security problems in uplink grant-free communications. To that end, the pilot in my approach is designed to be extracted from changeable environment, referred to as EA pilot, to keep users' pilots updated and thus ensure secure grant-free transmission. In detail, if a user's amplitude of channel information at the p-th pilot is higher than a set threshold at current transmission, the values of this users' pilot at the p-th pilot for the next transmission are set to 1, otherwise it is set to -1. In particular, the channel reciprocity in TDD mode is utilized to keep users' pilots changing without signaling exchange between users and BS. By doing so, the transmitter can determine the downlink channel information upon the channel information obtained through the uplink channel estimation. Hence, it is hard for eavesdroppers to acquire pilot information and steal information generated by the target users.

Since the value of pilot is set to 1 or -1, it can be observed that the values in $\hat{\boldsymbol{\alpha}}_{k,m}$ and $\bar{\boldsymbol{\alpha}}_{m,p}$ are random samples from a Bernoulli distribution and straightforward to implement. However, the pilot matrix should meet the requirement of the RIP to ensure the decoding at the BS [102, 103]. To this end, Bernoulli matrix is resorted to construct measurement

matrix, *i.e.,* random matrix. In fact, the measurement matrix in the time domain and angular domain can be respectively expressed as

$$\hat{\boldsymbol{\alpha}}_{k,r} = [\mathrm{diag}(\hat{\boldsymbol{\alpha}}_{k,r,0}), \ldots, \mathrm{diag}(\hat{\boldsymbol{\alpha}}_{k,r,N-1})], \tag{6.10}$$

and

$$\begin{aligned} \boldsymbol{\Phi}_p &= [\mathbf{A}_{\mathrm{BS}}^{\mathrm{T}} \otimes \bar{\boldsymbol{\alpha}}_{0,p}, \ldots, \mathbf{A}_{\mathrm{BS}}^{\mathrm{T}} \otimes \bar{\boldsymbol{\alpha}}_{N-1,p}] \\ &= [\boldsymbol{\Phi}_{0,p}, \ldots, \boldsymbol{\Phi}_{N-1,p}], \end{aligned} \tag{6.11}$$

where the element of $\hat{\boldsymbol{\alpha}}_{k,r,n}$ and $\bar{\boldsymbol{\alpha}}_{n,p}$ are generated by the independent and identically distributed Bernoulli random process. Accordingly, the pilot sequence $\boldsymbol{\alpha}$ is non-orthogonal. The proof of the validity of the measurement matrix is shown in Appendix B.

6.4 Simulation Results

In this section, the efficacy of the proposed scheme with EA pilots in LTE-A systems is demonstrated. Furthermore, the NMSE performance of eavesdroppers without knowledge of EA pilots is given to show the superiority of the proposed secure grant-free communications. The pilots are uniformly distributed throughout the symbol. Note that the linear interpolation algorithm is used to obtain channel information in non-pilot positions.

Considering the property of multipath slowly varying channels, the slight Doppler shift is employed. For example, given Doppler shift 5 Hz and the conventional Rayleigh fading channel, the coherence time T_c is equal to 8.46 ms based on (6.6), which is far less than a transmission subframe in LTE-A. Refer to [104], the channel information performance of the active users is evaluated based on the Extended Vehicular A model (EVA) channel model, where the multi-path propagation conditions consist of the maximum delay and relative power. The simulation parameters are given in Table 6.1.

Figure 6.3 shows impact of SNR for the proposed scheme on the NMSE performance of channel estimation in the time and the virtual angular domain, respectively. Furthermore, the NMSE performance at the eavesdropper side is presented to show the superiority of the EA pilots. It can be observed that by employing CS on the time sparse basis and virtual angular basis, the EA pilot approach brings high security in terms of NMSE of channel estimation, which means the eavesdropper is hard to perform channel estimation. Hence, it can be reasonably inferred that by utilizing the EA pilot in TDD mode, the uplink grant-free transmission can be conducted in a secure manner. In particular, it can

be seen that the performance of the channel estimation in the time domain outperforms the receiver in the angular virtual domain. This result is considered reasonable since the receiver with a small number of BS antennas is influenced by the leakage power.

Table 6.1: Simulation parameters for EA-based channel estimation

Parameters	**Values**
Mode	TDD
Frequency	2 GHz
Bandwith	20 MHz
Cyclic Prefix	Normal
Modulation Mode	QPSK
Channel Model	EVA
Doppler Shift	5 Hz
MIMO Correlation Matrices	Low
Number of the BS Antennas	64
Number of the Cluster	5
Number of Users in Cluster	2
Filter Coefficient	$\frac{1}{8}$

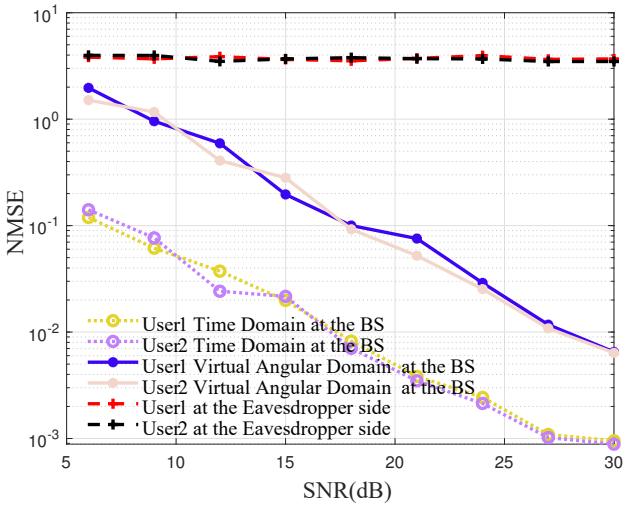

Figure 6.3: Impact of SNR on the NMSE performance of the proposed EA-based channel estimation in the time and virtual angular domain, respectively. Furthermore, the number of BS antennas is set to be 64.

6.5 Summary

CS based grant-free communication proposed in Chapter 3, 4 and 5 rely on the specific pilot matrix, which may be analyzed by an eavesdropper (Eve) to infer the user's channel information. In this context, the pilot design should be considered for secure grant-free communication. In this chapter, the standard CS form for channel estimation based on different sparse basis is derived. In particular, the essentially invariant character of the channel impulse within the coherence time has been utilized to perform channel estimation in the virtual angular domain. Furthermore, an EA pilot scheme is proposed to meet the requirement of secure grant-free communication, and the validity of the pilot design, *i.e.,* measurement matrix, is proven. Subsequently, the performance of the channel estimation based on different sparse basis is compared in the LTE-A system. The simulation results show that the proposed EA-pilots based grant-free communication is conducted in a secure manner.

Motivated by the requirement of CS based grant-free communication, UAD for aperiodic service, periodic service, the massive connected scenario and channel estimation for secure grant-free communication have been investigated in Chapter 3, 4, 5 and 6, respectively. In the following chapter, the conclusion and future work of this thesis are presented.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this thesis, CS based grant-free communication in the presence of massive potentially active users without moving or in low-speed moving scenarios is investigated. The main conclusions and findings are summarised as follows.

- Conventional request-grant based NOMA incurs tremendous overhead and high latency. To enable grant-free access in NOMA systems, UAD is essential. In this thesis, CS aided UAD is investigated, by utilizing the property of quasi-time-invariant channel tap delays as the prior information. This does not require any prior knowledge of the number of active users like the previous approaches, and therefore is more practical. Two UAD algorithms are proposed, which are referred to as g-TIDCS and mean value based and m-TIDCS, respectively. They achieve much higher UAD accuracy than the previous work at low SNR. Based on the UAD results, a low-complexity CS based channel estimation scheme is also proposed, which achieves higher accuracy than the previous channel estimation approaches.

- The grant-free communication with adaptive period for industrial IoT is investigated, where only a fraction of users are active at a time. To the best of my knowledge, this is the first work to exploit the non-continuous temporal correlation of the received signal for joint UAD, channel estimation and signal detection, while all the previous work requires continuous transmission. Two schemes are proposed toward this purpose, namely PBOMP and PBSBL, which outperform the previous schemes in terms of the

success rate of UAD, bit error rate and accuracy in period estimation and channel estimation. The CRLBs of channel estimation by PBOMP and PBSBL are derived. It is shown that the two proposed approaches have close CRLBs and normalized mean square error at high SNR.

- UAD and symbol detection for grant-free communication in the presence of massive users that are actively connected to base station (BS), where a small portion of to-be-connected users wake up in a burst. The number of effective interfering connected users is reduced to only one, by applying a preconditioning matrix to the received signals from multiple antennas at BS. Subsequently, an iterative UAD and SD scheme is applied, where the *priori* information about the remaining interfering user is exploited and the symbols of the burst users as well as the signature connected user are simultaneously detected by iterative exchanging of the soft information of user activity and symbols. The proposed system outperforms the existing work in terms of the success rate of UAD and bit error rate.

- CS has been extensively employed in grant-free communication, where data generated from different active users is transmitted to a BS without following the strict access grant process. Nevertheless, the state of the art CS algorithms rely on a very limited category of measurement matrix, *i.e.,* pilot matrix, which may be analyzed by an Eve to infer the user's channel information. Thus, the physical layer security becomes a critical issue in grant-free communication. In this thesis, the channel reciprocity in time-division duplex systems is utilized to design EA pilots derived from transmission channels to prevent eavesdroppers from acquiring users' channel information. The simulation results show that the proposed EA based pilot approach possesses a high level of security by scrambling the Eve' normalized mean square error performance of channel estimation.

7.2 Future Work

The work presented in this thesis could be extended in the aspects of multi-cluster coordination as shown in Subsection 7.2.1. Furthermore, utilizing interference for secure downlink transmission is also considered to construct a whole communication system with uplink transmission as shown in Subsection 7.2.2.

7.2.1 Joint User Detection and Channel Estimation for Grant-Free Multi-Cluster Coordination

IoT, capable of supporting massive access, sensing and interacting with everyone and everything, has been envisioned as a potential support for the imminent industry 4.0 [105]. The ubiquitous connectivity, which includes a variety of industrial users, such as robots, sensors, co-robots and other facilities, enables data collection, exchange and analysis. This potentially facilitates improvements in efficiency, productivity and other economic benefits. IoT networks aim to provide heterogeneous services for the massive users, where the use cases can be roughly classified into MMTC and URLLC. Specifically, MMTC, which is the focus of this subsection, is featured by massive connectivity, low power consumption and high energy efficiency, while URLLC is mainly designed to provide high reliability and low latency for serving mission critical applications, such as industrial real-time automation. Evidently, wireless communications in IoT cause severe access congestion due to insufficient frequency, time, power, antennas, and other resources. One of the efficient techniques is to cluster massive users into a number of small groups [106], with a part of resources being shared and reused in different clusters, referred to as multi-cluster IoT. Each cluster is equipped with one or more APs to provide high-quality wireless transmissions. A number of APs are connected by a high-speed optical fiber to data collection module and application server for resource coordination, authentication and interference management.

The wireless interface management between APs and users is essential in IoT, from the perspectives of radio spectrum allocation and interference management. The conventional radio resource management techniques have been extensively investigated in cellular networks, where frequency reuse is adopted among base stations. Nevertheless, the number of connections in IoT overwhelms that in cellular networks, typically requiring a dense AP deployment. The frequency reuse design leads to a low level of resource utilization and incurs stronger inter-cluster interference [107, 108], which is even severer for the users at the edge of the clusters, and significantly deteriorates their QoS. Considering the infrastructure of multi-cluster IoT where the deployed APs are connected to a centralized cloud manager for signal processing, the coordination techniques are naturally suitable for the wireless interface management in IoT. That is, multiple APs can coordinate with each other to share CSI and/or the intended transmission data for joint signal processing and interference control [105, 106]. More specifically, in [105], a coordination scheme was adopted in IoT to assist data offloading and assure load balance, while an energy-efficient coordina-

tion IoT framework was presented for balancing the traffic load, alongside an associated switching-on/off scheme for prolonging the lifetime of the system in [106]. Nevertheless, both the coordination designs in [105, 106] are based on the orthogonal multiuser access, and thus extra resources, *i.e.*, time, frequency and antennas, are required for multiuser access.

Low latency is an important metric for some IoT applications, such as early warning of malfunctioned equipments. However, the complicated handshake mechanism, that has been extensively used for grant access in cellular networks, leading to high access delay and outdated CSI. Considering the sporadic transmission at uplink, joint active devices detection and channel estimation would benefit system performance in terms of low access delay and accurate CSI acquiring. Since in practice the number of the active users may be much lower than the total number of users, it is reasonable to apply compressive sensing theory for UAD based on the transmission sparsity. The integration of NOMA and grant-free has been extensively researched, however, how to further improve the detection performance with the multicluster coordination needs more fundamental analysis.

7.2.2 Interference Utilization-based Precoder for Secure Communications

Downlink precoding has been regarded as a key technology in multiuser MIMO communications. With the CSI available at the base station, the multiuser interference can be calculated prior to transmission. In this way, the interference mitigation (IM)-based precoder techniques have been extensively investigated to strictly suppress the interference. The dirty-paper coding (DPC) scheme was proposed in [109] by pre-subtracting the interference prior to transmission for achieving capacity, which however assumes infinite alphabet input and incurs high computational cost. Although the Tomlinson-Harashima precoding (THP) [110] and vector perturbation (VP) [111] precoders aim to reduce the computational complexity over the DPC approach, they still need a sophisticated sphere-search algorithm for algorithm implementation. Hence, low-complexity linear precoders, such as zero-forcing (ZF) and MMSE [31, 112], have attracted much attention in practices due to their low-complexity. On the other hand, optimization-based precoding has been a popular research topic. For example, SNR balancing aims to maximize the minimum SNR subject to a total power constraint [113]; transmission power minimization problem aims to reduce the transmission power at the base station, subject to users' minimum SNR

requirements [114].

The above designs treat the input as infinite Gaussian signal. Hence, they only exploit the channel correlation for the precoding design. In practice, modulation size is finite, and the input is not Gaussian signal. In this case, there is scope to jointly exploit the correlation among the channels and transmitted data, so that the multiuser interference is possible to make constructive at each receiver, termed as interference utilization (IE) precoding [115]. The concept of constructive interference (CI) has been applied for anonymous communications [5], cognitive radio [116], large-scale MIMO [117, 118], constant envelope [119, 120], hybrid beamforming [121], multi-cell coordination [122], rate-splitting [123], physical layer security [1, 124, 125], directional modulation [6], and integrated sensing and communication systems [126].

The essential feature of future communications is that of supporting massive access in IoT, and therefore, the privacy and security requirements are intended to be more complicated and diversified due to the limited number of physical resources [55]. For example, the public broadcast may have a low privacy requirement, while some personal information requires high confidentiality [127]. A possible solution is to classify security rank and employ appropriate techniques of PHY security to meet the customized demand of different users. Hence, it is demanding to fundamental analysis and new metrics for designing and evaluating the overall system PHY security performance, especially under the perspective of the IE-based secure Communications [128].

Appendix A

A.1 Proof of (5.8) and (5.9) in *Lemma 5.3.2*

$L_{E,n}(\tilde{\mathbf{y}})$ in (5.7) is given by,

$$\begin{aligned} L_{E,n}(\tilde{\mathbf{y}}) &= \ln \frac{P(\tilde{\mathbf{y}}|x_n \in \varsigma, x_{m^*} \in \varsigma)}{P(\tilde{\mathbf{y}}|x_n = 0, x_{m^*} \in \varsigma)} \\ &= \ln \sum_{\varsigma_j} \exp(\ln \frac{P(\tilde{\mathbf{y}}|x_n = \varsigma_j, x_{m^*} \in \varsigma)}{P(\tilde{\mathbf{y}}|x_n = 0, x_{m^*} \in \varsigma)}) \\ &= \ln \sum_{\varsigma_j} \exp\ (L_{E,n,j}), \end{aligned} \tag{A.1}$$

Similarly, $L_{A,n}$ in (5.7) is given by,

$$\begin{aligned} L_{A,n} &= \ln \frac{P(x_n \in \varsigma, x_{m^*} \in \varsigma)}{P(x_n = 0, x_{m^*} \in \varsigma)} \\ &= \ln \sum_{\varsigma_j} \exp(\ln \frac{P(x_n = \varsigma_j, x_{m^*} = \varsigma)}{P(x_n = 0, x_{m^*} \in \varsigma)}) \\ &= \ln \sum_{\varsigma_j} \exp\ (L_{A,n,j}). \end{aligned} \tag{A.2}$$

A.2 Proof of (5.10)

The index of the burst user is in line with the maximum probability of the *posteriori* LLR of user activity, we have

$$\begin{aligned} n^* &= \arg\max_{n\in\bar{\mathbb{S}}^{(l-1)}} L_n(\tilde{\mathbf{y}}) \\ &= \arg\max_{n\in\bar{\mathbb{S}}^{(l-1)}} (\ln \sum_{\varsigma_j} \exp(L_{E_1,n,j}(\tilde{\mathbf{y}}) + L_{A,n,j})) \\ &\overset{(a)}{=} \arg\max_{n\in\bar{\mathbb{S}}^{(l-1)}} (\ln \sum_{\varsigma_j} \exp(L_{E_1,n,j}(\tilde{\mathbf{y}})) + L_{A,n}) \\ &\approx \arg\max_{n\in\bar{\mathbb{S}}^{(l-1)}} (\max_{j} L_{E_1,n,j}(\tilde{\mathbf{y}}) + L_{A,n}), \end{aligned} \tag{A.3}$$

where (a) is due to the fact that the elements ς_j of ς are active with equal probability and constant is omitted for simplicity. n^* is selected as the index of the burst users.

A.3 Proof of (5.14)

The mathematical expectation of the data generated by the k-th potential active user is calculated as

$$\begin{aligned} \bar{x}_k &= \sum_{\varsigma_j\in\varsigma} \mathrm{P}\left(x_k = \varsigma_j\right)\varsigma_j \\ &= \sum_{\varsigma_j\in\varsigma} \frac{\exp(L_{A_1,k,j})\varsigma_j}{1+\exp(L_{A_1,k,j})} \\ &\overset{(b)}{=} \frac{\exp\left(L_{A,k} - \ln|\varsigma|\right) \sum\limits_{\varsigma_j\in\varsigma} \exp(L_{E_2,k,j})\varsigma_j}{1+\exp\left(L_{A,k} - \ln|\varsigma|\right) \sum\limits_{\varsigma_j\in\zeta} \exp(L_{E_2,k,j})}, \end{aligned} \tag{A.4}$$

where $L_{A_1,k,j}$ indicates *a priori* information of active symbol at the l-th iteration. $L_{A_1,k,j}$ at the l-th iteration is divided into two parts to obtain the soft symbol information at the (l-1)-th iteration [68]. Explicitly, $L_{A_1,k,j} = L_{A,k,j} + L_{E_2,k,j}$, where $L_{A,k,j}$ is in line with original *priori* information of active symbol. As mentioned early, the elements ς_j of ς are active with equal probability, and hence, $L_{A,k,j} = L_{A,k} - \ln|\varsigma|$. Given this guideline, (b) in (A.4) can be obtained.

Appendix B

The authors in [129] utilized statistical RIP to construct measurement matrix, where the measurement matrix fulfill (B.1)-(B.4). Given this guideline, $\hat{\boldsymbol{\alpha}}_{k,r}$ and $\boldsymbol{\Phi}_p$ is proved to can be used to design measurement matrix based on the different sparse domain, respectively. For simplicity, $\boldsymbol{\alpha}$ is employed to denote $\hat{\boldsymbol{\alpha}}_{k,r}$ in subsection B.1 and $\boldsymbol{\Phi}_p$ in subsection B.2.

B.1 Proof of the measurement matrix in the time domain

- Since the pilots follow Bernoulli random distribution, it can be concluded that,

$$\sum_{j \neq i} \boldsymbol{\alpha}_j \boldsymbol{\alpha}_i^{\mathrm{T}} = 0, \tag{B.1}$$

$$\sum_{cl=0}^{PM} \boldsymbol{\alpha}_j(cl) = 0, \tag{B.2}$$

  where $\boldsymbol{\alpha}_j$ and $\boldsymbol{\alpha}_i$ denote the j-th and i-th column of $\boldsymbol{\alpha}$, respectively.

- The columns of form a group under "pointwise multiplication",

$$\begin{aligned} &\forall\ column\ j^{''}, \exists\ j \neq j^{'} \neq j^{''}, \\ &\boldsymbol{\alpha}_{j^{''}}(cl) = \boldsymbol{\alpha}_j(cl)\boldsymbol{\alpha}_{j^{'}}(cl). \end{aligned} \tag{B.3}$$

  Since non-diagonal element of $\hat{\boldsymbol{\alpha}}_{k,r,n}$ is zero, the diagonal element of $\hat{\boldsymbol{\alpha}}_{k,r,n}$ is considered. Now recall that the diagonal element of $\hat{\boldsymbol{\alpha}}_{k,r,n}$ is 1 or -1. Hence, (B.3) holds.

- For all $j \in 2, \ldots, PM$,

$$|\sum_{row=0}^{P-1} \boldsymbol{\alpha}_j(row)|^2 \leq P^{2-\eta}, 0 < \eta \leq 1. \tag{B.4}$$

B.2 Proof of the measurement matrix in the virtual angular domain

Similar to the proof of measurement matrix design in the time domain, (B.1) – (B.4) still holds in the virtual angular domain. To give the proof in the virtual angular domain and based on elementary column transformation, (6.11) is rewritten as

$$\begin{aligned}\bar{\bar{\boldsymbol{\Phi}}}_p =&[\bar{\alpha}_{0,0,p}\mathbf{A}_{\mathrm{BS}}^{\mathrm{T}}, \ldots, \bar{\alpha}_{K-1,0,p}\mathbf{A}_{\mathrm{BS}}^{\mathrm{T}}, \ldots, \\ &\bar{\alpha}_{0,N-1,p}\mathbf{A}_{\mathrm{BS}}^{\mathrm{T}}, \ldots, \bar{\alpha}_{K-1,N-1,p}\mathbf{A}_{\mathrm{BS}}^{\mathrm{T}}].\end{aligned} \tag{B.5}$$

Refer to (B.5), since $\mathbf{A}_{\mathrm{BS}}$ is a DFT matrix and the rows of $\bar{\bar{\boldsymbol{\Phi}}}_p$ are orthogonal each other. Hence, (B.1) holds. Furthermore, $\bar{\alpha}_{k,n,p}$ follows Bernoulli random distribution with 1 or -1, the sum of any rows is zero if the number of rows enough long. Hence, (B.2) can be obtained.

Since the block matrix of $\bar{\bar{\boldsymbol{\Phi}}}_p$ denotes a DFT matrix, and the element of the first column of $\mathbf{A_{BS}}$ is set to 1. Hence, it can be always found different rows satisfy (B.3). In this context, the maximum of element of $\bar{\bar{\boldsymbol{\Phi}}}_p$ is equal to 1, and therefore, (B.4) holds.

Bibliography

[1] Z. Wei, C. Masouros, F. Liu, S. Chatzinotas, and B. Ottersten, "Energy- and Cost-Efficient Physical Layer Security in the Era of IoT: The Role of Interference," *IEEE Communications Magazine*, vol. 58, no. 4, pp. 81–87, 2020.

[2] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. D. Renzo, and M. Debbah, "Holographic MIMO Surfaces for 6G Wireless Networks: Opportunities, Challenges, and Trends," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 118–125, 2020.

[3] H. Xu, W. Yu, D. Griffith, and N. Golmie, "A Survey on Industrial Internet of Things: A Cyber-Physical Systems Perspective," *IEEE Access*, vol. 6, pp. 78238–78259, 2018.

[4] J. Cao, X. Zhu, Y. Jiang, Z. Wei, and S. Sun, "Information Age-Delay Correlation and Optimization With Finite Block Length," *IEEE Transactions on Communications*, vol. 69, no. 11, pp. 7236–7250, 2021.

[5] Z. Wei, F. Liu, C. Masouros, and H. V. Poor, "Fundamentals of Physical Layer Anonymous Communications: Sender Detection and Anonymous Precoding," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 64–79, 2022.

[6] Z. Wei, C. Masouros, and F. Liu, "Secure Directional Modulation With Few-Bit Phase Shifters: Optimal and Iterative-Closed-Form Designs," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 486–500, 2021.

[7] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[8] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive Access for 5G and Beyond," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 615–637, 2021.

[9] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, Opportunities, and Directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724–4734, 2018.

[10] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G Internet of Things: A Comprehensive Survey," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 359–383, 2022.

[11] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-Reliable and Low-Latency Communications in 5G Downlink: Physical Layer Aspects," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 124–130, 2018.

[12] M. Rahnema, "Overview of the GSM System and Protocol Architecture," *IEEE Communications Magazine*, vol. 31, no. 4, pp. 92–100, 1993.

[13] K. Gilhousen, I. Jacobs, R. Padovani, A. Viterbi, L. Weaver, and C. Wheatley, "On the Capacity of a Cellular CDMA System," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 2, pp. 303–312, 1991.

[14] F. Ghavimi and H.-H. Chen, "M2M Communications in 3GPP LTE/LTE-A Networks: Architectures, Service Requirements, Challenges, and Applications," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 2, pp. 525–549, 2015.

[15] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges," *IEEE Access*, vol. 6, pp. 3619–3647, 2018.

[16] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovacs, and T. K. Madsen, "System Level Analysis of Uplink Grant-Free Transmission for URLLC," in *IEEE Globecom Workshops (GC Wkshps)*, Singapore, Dec. 2017.

[17] K. Senel and E. G. Larsson, "Grant-Free Massive MTC-Enabled Massive MIMO: A Compressive Sensing Approach," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6164–6175, 2018.

[18] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse Signal Processing for Grant-Free Massive Connectivity: A Future Paradigm for Random Access Protocols in the Internet of Things," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 88–99, 2018.

[19] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, "Analyzing Grant-Free Access for URLLC Service," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 741–755, 2021.

[20] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 4, pp. 3098–3130, 2018.

[21] A. C. Cirik, N. M. Balasubramanya, L. Lampe, G. Vos, and S. Bennett, "Toward the Standardization of Grant-Free Operation and the Associated NOMA Strategies in 3GPP," *IEEE Communications Standards Magazine*, vol. 3, no. 4, pp. 60–66, 2019.

[22] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-Free Non-Orthogonal Multiple Access for IoT: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 3, pp. 1805–1838, 2020.

[23] Y. Huang, C. Zhang, J. Wang, Y. Jing, L. Yang, and X. You, "Signal Processing for MIMO-NOMA: Present and Future Challenges," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 32–38, 2018.

[24] M. Baghani, S. Parsaeefard, M. Derakhshani, and W. Saad, "Dynamic Non-Orthogonal Multiple Access and Orthogonal Multiple Access in 5G Wireless Networks," *IEEE Transactions on Communications*, vol. 67, no. 9, pp. 6360–6373, 2019.

[25] J.-B. Seo, B. C. Jung, and H. Jin, "Nonorthogonal Random Access for 5G Mobile Communication Systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7867–7871, 2018.

[26] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of Things in the 5G Era: Enablers, Architecture, and Business Models," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 510–527, 2016.

[27] X. Liu, B. Lin, M. Zhou, and M. Jia, "NOMA-Based Cognitive Spectrum Access for 5G-Enabled Internet of Things," *IEEE Network*, vol. 35, no. 5, pp. 290–297, 2021.

[28] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of Non-Orthogonal Multiple Access in LTE and 5G Networks," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 185–191, 2017.

[29] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive Non-Orthogonal Multiple Access for Cellular IoT: Potentials and Limitations," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 55–61, 2017.

[30] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "A Novel Analytical Framework for Massive Grant-Free NOMA," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2436–2449, 2019.

[31] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge university press, 2005.

[32] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and Practical Considerations of Non-Orthogonal Multiple Access (NOMA) for Future Radio Access," in *International Symposium on Intelligent Signal Processing and Communication Systems*, Naha, Japan, Nov. 2013, pp: 770-774.

[33] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-s. Kwak, "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 2, pp. 721–742, 2017.

[34] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, "Uplink Non-Orthogonal Multiple Access for 5G Wireless Networks," in *11th International Symposium on Wireless Communications Systems (ISWCS)*, Barcelona, Spain, Aug. 2014, pp: 781-785.

[35] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel Low-Density Signature for Synchronous CDMA Systems Over AWGN Channel," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1616–1626, 2008.

[36] H. Nikopour, E. Yi, A. Bayesteh, K. Au, M. Hawryluck, H. Baligh, and J. Ma, "SCMA for Downlink Multiple Access of 5G Wireless Networks," in *IEEE Global Communications Conference (GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp: 3940-3945.

[37] T. Abe and T. Matsumoto, "Space-time Turbo Equalization in Frequency-Selective MIMO Channels," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 3, pp. 469–475, 2003.

[38] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern Division Multiple Access—A Novel Nonorthogonal Multiple Access for Fifth-Generation Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, 2017.

[39] Q. Wang, R. Zhang, L.-L. Yang, and L. Hanzo, "Non-Orthogonal Multiple Access: A Unified Perspective," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 10–16, 2018.

[40] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO Detection Techniques: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019.

[41] Z. Qin, J. Fan, Y. Liu, Y. Gao, and G. Y. Li, "Sparse Representation for Wireless Communications: A Compressive Sensing Approach," *IEEE Signal Processing Magazine*, vol. 35, no. 3, pp. 40–58, 2018.

[42] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[43] E. Crespo Marques, N. Maciel, L. Naviner, H. Cai, and J. Yang, "A Review of Sparse Recovery Algorithms," *IEEE Access*, vol. 7, pp. 1300–1322, 2019.

[44] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal Processing With Compressive Measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, 2010.

[45] K. Yang, N. Yang, N. Ye, M. Jia, Z. Gao, and R. Fan, "Non-Orthogonal Multiple Access: Achieving Sustainable Future Radio Access," *IEEE Communications Magazine*, vol. 57, no. 2, pp. 116–121, 2019.

[46] H. Zeng, X. Zhu, Y. Jiang, Z. Wei, and T. Wang, "A Green Coordinated Multi-Cell NOMA System With Fuzzy Logic Based Multi-Criterion User Mode Selection and Resource Allocation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 480–495, 2019.

[47] B. Wang, L. Dai, Y. Zhang, T. Mir, and J. Li, "Dynamic Compressive Sensing-Based Multi-User Detection for Uplink Grant-Free NOMA," *IEEE Communications Letters*, vol. 20, no. 11, pp. 2320–2323, 2016.

[48] Y. Du, C. Cheng, B. Dong, Z. Chen, X. Wang, J. Fang, and S. Li, "Block-Sparsity-Based Multiuser Detection for Uplink Grant-Free NOMA," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 7894–7909, 2018.

[49] J. Xiao, G. Deng, G. Nie, H. Tian, and J. Jin, "Dynamic Adaptive Compressive Sensing-Based Multi-User Detection in Uplink URLLC," in *IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Bologna, Italy, Sept. 2018, pp: 1-6.

[50] L. Qiao, J. Zhang, Z. Gao, S. Chen, and L. Hanzo, "Compressive Sensing Based Massive Access for IoT Relying on Media Modulation Aided Machine Type Communications," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 10391–10396, 2020.

[51] Y. Zhang, Q. Guo, Z. Wang, J. Xi, and N. Wu, "Block Sparse Bayesian Learning Based Joint User Activity Detection and Channel Estimation for Grant-Free NOMA Systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9631–9640, 2018.

[52] B. Wang, L. Dai, T. Mir, and Z. Wang, "Joint User Activity and Data Detection Based on Structured Compressive Sensing for NOMA," *IEEE Communications Letters*, vol. 20, no. 7, pp. 1473–1476, 2016.

[53] Y. Nan, L. Zhang, and X. Sun, "Efficient Downlink Channel Estimation Scheme Based on Block-Structured Compressive Sensing for TDD Massive MU-MIMO Systems," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 345–348, 2015.

[54] M. Alam and Q. Zhang, "Sequence Block Based Compressed Sensing Multiuser Detection for 5G," in *IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp: 1-6.

[55] Y. Wang, X. Zhu, E. G. Lim, Z. Wei, Y. Liu, and Y. Jiang, "Compressive Sensing based User Activity Detection and Channel Estimation in Uplink NOMA Sys-

tems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, South Korea, May 2020, pp: 1-6.

[56] Q.-Y. Yu, H.-C. Lin, and H.-H. Chen, "Intelligent Radio for Next Generation Wireless Communications: An Overview," *IEEE Wireless Communications*, vol. 26, no. 4, pp. 94–101, 2019.

[57] Y. Yuan, Z. Yuan, and L. Tian, "5G Non-Orthogonal Multiple Access Study in 3GPP," *IEEE Communications Magazine*, vol. 58, no. 7, pp. 90–96, 2020.

[58] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of Compressive Sensing to Sparse Channel Estimation," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 164–174, 2010.

[59] Z. Gao, L. Dai, S. Han, C.-L. I, Z. Wang, and L. Hanzo, "Compressive Sensing Techniques for Next-Generation Wireless Communications," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 144–153, 2018.

[60] R. T. Yazicigil, T. Haque, P. R. Kinget, and J. Wright, "Taking Compressive Sensing to the Hardware Level: Breaking Fundamental Radio-Frequency Hardware Performance Tradeoffs," *IEEE Signal Processing Magazine*, vol. 36, no. 2, pp. 81–100, 2019.

[61] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative Channel Estimation Using LSE and Sparse Message Passing for MmWave MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 245–259, 2019.

[62] B. Shim and B. Song, "Multiuser Detection via Compressive Sensing," *IEEE Communications Letters*, vol. 16, no. 7, pp. 972–974, 2012.

[63] X. Zhang, W. Xu, Y. Cui, L. Lu, and J. Lin, "On Recovery of Block Sparse Signals via Block Compressive Sampling Matching Pursuit," *IEEE Access*, vol. 7, pp. 175554–175563, 2019.

[64] S. H. Fouladi, S.-E. Chiu, B. D. Rao, and I. Balasingham, "Recovery of Independent Sparse Sources From Linear Mixtures Using Sparse Bayesian Learning," *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6332–6346, 2018.

[65] 3GPP, "Service Requirements for Cyber-Physical Control Applications in Vertical Domains; stage-1(Release 17)," TS 22.104, the 3rd Generation Partnership Project (3GPP), Jun. 2019.

[66] B. Gedik, L. Liu, and P. S. Yu, "ASAP: An Adaptive Sampling Approach to Data Collection in Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 12, pp. 1766–1783, 2007.

[67] J.-B. Seo, "On Minimizing Energy Consumption of Duty-Cycled Wireless Sensors," *IEEE Communications Letters*, vol. 19, no. 10, pp. 1698–1701, 2015.

[68] B. K. Jeong, B. Shim, and K. B. Lee, "MAP-Based Active User and Data Detection for Massive Machine-Type Communications," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8481–8494, 2018.

[69] Y. Wang, X. Zhu, E. G. Lim, Z. Wei, and Y. Jiang, "Grant-Free Communications With Adaptive Period for IIoT: Sparsity and Correlation-Based Joint Channel Estimation and Signal Detection," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4624–4638, 2022.

[70] X. Wu, G. Yang, F. Hou, and S. Ma, "Low-Complexity Downlink Channel Estimation for Millimeter-Wave FDD Massive MIMO Systems," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1103–1107, 2019.

[71] S. Li, L. Xiao, and T. Jiang, "An Efficient Matching Pursuit Based Compressive Sensing Detector for Uplink Grant-Free NOMA," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 2012–2017, 2021.

[72] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6169–6183, 2015.

[73] J. Dai, A. Liu, and V. K. N. Lau, "FDD Massive MIMO Channel Estimation With Arbitrary 2D-Array Geometry," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2584–2599, 2018.

[74] Y. Du, B. Dong, W. Zhu, P. Gao, Z. Chen, X. Wang, and J. Fang, "Joint Channel Estimation and Multiuser Detection for Uplink Grant-Free NOMA," *IEEE Wireless Communications Letters*, vol. 7, no. 4, pp. 682–685, 2018.

[75] Y. Rachlin and D. Baron, "The Secrecy of Compressed Sensing Measurements," in *46th Annual Allerton conference on Communication, Control, and Computing*, IEEE, Monticello, IL, USA, Sept. 2008, pp: 813-817.

[76] X. Zhu, L. Dai, W. Dai, Z. Wang, and M. Moonen, "Tracking a Dynamic Sparse Channel via Differential Orthogonal Matching Pursuit," in *MILCOM 2015 - 2015 IEEE Military Communications Conference*, pp. 792–797, 2015.

[77] W. Shen, L. Dai, Y. Shi, X. Zhu, and Z. Wang, "Compressive Sensing-based Differential Channel Feedback for Massive MIMO," *Electronics Letters*, vol. 51, no. 22, pp. 1824–1826, 2015.

[78] F. Monsees, M. Woltering, C. Bockelmann, and A. Dekorsy, "Compressive Sensing Multi-User Detection for Multicarrier Systems in Sporadic Machine Type Communication," in *IEEE 81st Vehicular Technology Conference (VTC Spring)*, Glasgow, UK, May 2015, pp: 1-5.

[79] D. Needell and J. A. Tropp, "CoSaMP: Iterative Signal Recovery From Incomplete and Inaccurate Samples," *Applied and computational harmonic analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[80] J. A. Tropp and A. C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[81] J. Zhang, L. Lu, Y. Sun, Y. Chen, J. Liang, J. Liu, H. Yang, S. Xing, Y. Wu, J. Ma, I. B. F. Murias, and F. J. L. Hernando, "PoC of SCMA-Based Uplink Grant-Free Transmission in UCNC for 5G," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1353–1362, 2017.

[82] L. Wu, P. Sun, Z. Wang, and Y. Yang, "Joint User Activity Identification and Channel Estimation for Grant-Free NOMA: A Spatial–Temporal Structure-Enhanced Approach," *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12339–12349, 2021.

[83] A. Zaidi, F. Athley, J. Medbo, U. Gustavsson, G. Durisi, and X. Chen, *5G Physical Layer: Principles, Models and Technology Components*. Academic Press, 2018.

[84] Mousavi, Ahmad and Rezaee, Mehdi and Ayanzadeh, Ramin, "A Survey on Compressive Sensing: Classical Results and Recent Advancements," *arXiv preprint arXiv:1908.01014*, 2019.

[85] Z. Zhang and B. D. Rao, "Extension of SBL Algorithms for the Recovery of Block Sparse Signals With Intra-Block Correlation," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2009–2015, 2013.

[86] Z. Zhang and B. D. Rao, "Sparse Signal Recovery With Temporally Correlated Source Vectors Using Sparse Bayesian Learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 912–926, 2011.

[87] M. R. O'Shaughnessy, M. A. Davenport, and C. J. Rozell, "Sparse Bayesian Learning With Dynamic Filtering for Inference of Time-Varying Sparse Signals," *IEEE Transactions on Signal Processing*, vol. 68, pp. 388–403, 2020.

[88] S. Ji, Y. Xue, and L. Carin, "Bayesian Compressive Sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[89] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of machine learning research*, vol. 1, no. Jun, pp. 211–244, 2001.

[90] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[91] S. M. Kay, *Fundamentals of Statistical Signal Srocessing*. Prentice Hall PTR, 1993.

[92] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive Sensing-Based Adaptive Active User Detection and Channel Estimation: Massive Access Meets Massive MIMO," *IEEE Transactions on Vehicular Technology*, vol. 68, pp. 764–779, 2020.

[93] H. Li and J. Wen, "A New Analysis for Support Recovery With Block Orthogonal Matching Pursuit," *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 247–251, 2019.

[94] Z. Zhang, Y. Li, C. Huang, Q. Guo, L. Liu, C. Yuen, and Y. L. Guan, "User Activity Detection and Channel Estimation for Grant-Free Random Access in LEO Satellite-Enabled Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8811–8825, 2020.

[95] Y. Wu, X. Gao, S. Zhou, W. Yang, Y. Polyanskiy, and G. Caire, "Massive Access for Future Wireless Communication Systems," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 148–156, 2020.

[96] L. Wu, Z. Zhang, J. Dang, Y. Wu, H. Liu, and J. Wang, "Joint User Identification and Channel Estimation Over Rician Fading Channels," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6803–6807, 2020.

[97] Z. Zhang, Y. Li, C. Huang, Q. Guo, C. Yuen, and Y. L. Guan, "DNN-Aided Block Sparse Bayesian Learning for User Activity Detection and Channel Estimation in Grant-Free Non-Orthogonal Random Access," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12000–12012, 2019.

[98] L.-U. Choi and R. Murch, "A Transmit Preprocessing Technique for Multiuser MIMO Systems Using a Decomposition Approach," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 20–24, 2004.

[99] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A Survey of Sparse Representation: Algorithms and Applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.

[100] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental Study of MIMO Channel Statistics and Capacity via the Virtual Channel Representation," *Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep*, vol. 5, pp. 10–15, 2007.

[101] W. C. Lee, *Mobile Communications Engineering: Theory and Applications*. McGraw-Hill Education, 1998.

[102] O. Scherzer, *Handbook of Mathematical Methods in Imaging*. Springer Science & Business Media, 2010.

[103] Y. Ding, D. Kunisky, A. S. Wein, and A. S. Bandeira, "The Average-Case Time Complexity of Certifying the Restricted Isometry Property," *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7355–7361, 2021.

[104] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception," TS 36.101, the 3rd Generation Partnership Project (3GPP), sep 2018.

[105] T. Qiu, J. Chi, X. Zhou, Z. Ning, M. Atiquzzaman, and D. O. Wu, "Edge Computing in Industrial Internet of Things: Architecture, Advances and Challenges," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 4, pp. 2462–2488, 2020.

[106] K. Wang, Y. Wang, Y. Sun, S. Guo, and J. Wu, "Green Industrial Internet of Things Architecture: An Energy-Efficient Perspective," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 48–54, 2016.

[107] J. An, K. Wang, S. Wang, Y. Lin, and X. Bu, "Antenna Array Calibration for IIoT Oriented Satellites: From Orthogonal CDMA to NOMA," *IEEE Wireless Communications*, vol. 27, no. 6, pp. 28–36, 2020.

[108] B. Selim, M. S. Alam, J. V. C. Evangelista, G. Kaddoum, and B. L. Agba, "NOMA-Based IoT Networks: Impulsive Noise Effects and Mitigation," *IEEE Communications Magazine*, vol. 58, no. 11, pp. 69–75, 2020.

[109] M. Costa, "Writing on Dirty Paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[110] L. Sun and M. Lei, "Quantized CSI-Based Tomlinson-Harashima Precoding in Multiuser MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1118–1126, 2013.

[111] B. Hochwald, C. Peel, and A. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication-Part II: Perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537–544, 2005.

[112] C. Peel, B. Hochwald, and A. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication-Part I: Channel Inversion and Regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, 2005.

[113] A. Wiesel, Y. Eldar, and S. Shamai, "Linear Precoding via Conic Optimization for Fixed MIMO Receivers," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 161–176, 2006.

[114] M. Schubert and H. Boche, "Solution of the Multiuser Downlink Beamforming Problem with Individual SINR Constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, 2004.

[115] C. Masouros and E. Alsusa, "A Novel Transmitter-Based Selective-Precoding Technique for DS/CDMA Systems," *IEEE Signal Processing Letters*, vol. 14, no. 9, pp. 637–640, 2007.

[116] K. L. Law and C. Masouros, "Symbol Error Rate Minimization Precoding for Interference Exploitation," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5718–5731, 2018.

[117] P. V. Amadori and C. Masouros, "Interference-Driven Antenna Selection for Massive Multiuser MIMO," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 5944–5958, 2016.

[118] P. V. Amadori and C. Masouros, "Large Scale Antenna Selection and Precoding for Interference Exploitation," *IEEE Transactions on Communications*, vol. 65, no. 10, pp. 4529–4542, 2017.

[119] P. V. Amadori and C. Masouros, "Constant Envelope Precoding by Interference Exploitation in Phase Shift Keying-Modulated Multiuser Transmission," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 538–550, 2017.

[120] F. Liu, C. Masouros, P. V. Amadori, and H. Sun, "An Efficient Manifold Algorithm for Constructive Interference Based Constant Envelope Precoding," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1542–1546, 2017.

[121] G. Hegde, C. Masouros, and M. Pesavento, "Interference Exploitation-Based Hybrid Precoding With Robustness Against Phase Errors," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3683–3696, 2019.

[122] Z. Wei, C. Masouros, K.-K. Wong, and X. Kang, "Multi-Cell Interference Exploitation: Enhancing the Power Efficiency in Cell Coordination," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 547–562, 2020.

[123] A. Salem, C. Masouros, and K.-K. Wong, "Sum Rate and Fairness Analysis for the MU-MIMO Downlink Under PSK Signalling: Interference Suppression vs Exploitation," *IEEE Transactions on Communications*, vol. 67, no. 9, pp. 6085–6098, 2019.

[124] M. R. A. Khandaker, C. Masouros, K.-K. Wong, and S. Timotheou, "Secure SWIPT by Exploiting Constructive Interference and Artificial Noise," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1326–1340, 2019.

[125] Z. Wei and C. Masouros, "Device-Centric Distributed Antenna Transmission: Secure Precoding and Antenna Selection With Interference Exploitation," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2293–2308, 2020.

[126] F. Liu, C. Masouros, A. Li, T. Ratnarajah, and J. Zhou, "MIMO Radar and Cellular Coexistence: A Power-Efficient Approach Enabled by Interference Exploitation," *IEEE Transactions on Signal Processing*, vol. 66, no. 14, pp. 3681–3695, 2018.

[127] L. Chen, J. Li, and Y. Zhang, "Anonymous Certificate-Based Broadcast Encryption With Personalized Messages," *IEEE Transactions on Broadcasting*, vol. 66, no. 4, pp. 867–881, 2020.

[128] Z. Wei, C. Masouros, H. V. Poor, A. P. Petropulu, and L. Hanzo, "Physical Layer Anonymous Precoding: The Path to Privacy-Preserving Communications," *arXiv preprint arXiv:2109.08876*, 2021.

[129] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a Large Class of Deterministic Sensing Matrices That Satisfy a Statistical Isometry Property," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 358–374, 2010.