Grant-Free Non-Orthogonal Multiple Access for Massive Low-Latency Communications

Takanori Hara

Department of Computer and Network Engineering
The University of Electro-Communications

This dissertation is submitted in partial fulfillment of the requirements for the degree of *Doctor of Philosophy*

March 2022

Document Description

Title:

Grant-Free Non-Orthogonal Multiple Access for Massive Low-Latency Communications

Title (in Japanese):

大規模低遅延通信のためのグラントフリー非直交多元接続

Author:

Takanori Hara

Supervisory Committee

Chairperson: Professor Koji Ishibashi

1. Member: Professor Takeo Fujii

2. Member: Professor Kazunori Hayashi

3. Member: Associate Professor Hideki Yagi

4. Member: Associate Professor Koichi Adachi

Day of the Predefense: November 5th, 2021.
Day of the Defense: February 3rd, 2022.

Acknowledgements

First of all, I am sincerely grateful to my supervisor, Professor Koji Ishibashi, for his kind and ardent support. His dedicated guidance, attitude toward research, and thoughtful discussions have largely motivated me to focus on our research. Thanks to his strong support, I could have a lot of exciting and irreplaceable experiences in the last six years in our laboratory, making my life more colorful.

Undoubtedly, the experience in my visiting the Communications Group headed by Professor Lajos Hanzo has triggered me to go for the PhD course, and I would therefore like to express my gratitude to him. I would also like to thank the dissertation committee members, Professor Takeo Fujii, Professor Kazunori Hayashi, Associate Professor Hideki Yagi, and Associate Professor Koichi Adachi, for spending considerable time reading my dissertation and providing their insightful comments.

Joyously, during my PhD study, I have been able to collaborate with many researchers. My special thanks go to Professor Giuseppe Thadeu Freitas de Abreu, Dr. Razvan-Andrei Stoica, Dr. David González G., Dr. Osvaldo Gonsa, Dr. Omid Taghizadeh, and Hiroki Iimori, who have always shared insightful ideas and comments, leading to wide-ranging research outputs. Moreover, I am greatly benefited from Prof. Takeo Ohgane, Prof. Toshihiko Nishimura, Prof. Shinsuke Ibi, Assistant Prof. Ryo Hayakawa, Assistant Prof. Takumi Takahashi, and the members in their laboratory. Their insightful comments and suggestions greatly improved the quality of my research.

I am also grateful to Emeritus Professor Yasushi Yamao, the current and previous lab members, and my colleagues in Advanced Wireless Communication Research Center for encouraging and productive discussion.

Last but not least, my special thanks go to my family for their continuous help and support throughout my life. Without their continued patience and endless support, this dissertation would not have been possible.

Takanori Hara
Osaka, Japan
October 8, 2021

概要

自動運転やロボット制御などのアプリケーションでは多数の無線端末が，低遅延で通信する必要性がある．既存の無線通信システムではデータ送信の際に，通信の許可であるグラントを基地局から取得する必要があり，端末数の増加に伴って，通信遅延が増加してしまう．これを解決する手段として，各端末がグラント取得の手続きを踏まずに，即座にデータを伝送するグラントフリー伝送と，非直交多元接続を組み合わせたグラントフリー非直交多元接続（GF-NOMA）が注目されており，多数接続と低遅延性を同時に満足する技術として有望視されている．一方，基地局では，干渉信号から送信端末および送信データを効率的かつ高信頼に推定する必要がある．本研究では，これを解決する技術として圧縮センシングに着目し，多数接続と低遅延性を同時に満足するGF-NOMAの設計と，その効率的復調法を提案した．本論文の構成は以下の通りである．

第２章では，GF-NOMAの基本的な検討として，周波数非選択性通信路における時間領域拡散を用いたGF-NOMAを想定し，複数観測値を用いた近似メッセージ伝播法（MMV-AMP）に基づく手法を２つ提案した．１つは，MMV-AMPを直接用いて伝送を実際に行う端末（アクティブ端末），伝播路利得および送信データを一括して推定する手法を検討し，その見逃し確率特性について，状態発展法に基づいた理論解析を行うとともに，その解析に基づいたアクティブ端末の検出に用いる判定閾値の設計を行なった．計算機シミュレーションにより，理論的な結果と近い見逃し確率特性が得られることを示した．また，再設計した判定閾値を用いることで，アクティブ端末を既知とし，最小平均二乗誤差検出器を用いて伝播路および送信データを推定した場合のシンボル誤り率特性に漸近した特性を達成できることを明らかにした．もう１つの手法は，基地局で要する計算量の更なる削減を図るため，所望信号の次元を縮小する変換処理を施し，MMV-AMPによる推定を繰り返し行う．数値結果より，逆行列計算を必要とする従来手法よりも大幅に計算量を削減しつつ，同程度の伝播路および送信データの推定精度を達成できることを明らかにした．

次に，第３章では，時間領域拡散を用いたGF-NOMAの実装上の問題に対処するため，無線伝播路の事前情報を不要とした伝播路推定手法および各端末で生じる搬送波周波数オフセットを考慮した端末検出手法をそれぞれ提案した．前

者の手法では，MMV-AMPに期待値最大化アルゴリズムを導入することでMMV-AMPを拡張し，パスロスやシャドウイングなどを含む，伝播路の長区間変動に関する事前情報を必要としない高精度な伝播路推定を可能とする．後者の手法では，アレイ信号処理の考えに基づいた式変形を施した後，共分散推定における最尤推定問題としてアクティブ端末の検出問題を定式化し，座標降下法を用いて効率的に解く．計算機シミュレーションにより，先行研究と比較することで，それぞれの手法の有効性を示した．

時間領域拡散を用いたGF-NOMAでは，要求された遅延時間内に収めるために周波数帯域を広げて伝送速度を上げる必要がある．一方，広帯域化に伴って通信路が周波数選択性となるため，周波数非選択性通信路に対して設計された手法は適用できない．そこで，第４章では，通信路が周波数選択性である場合に，要求伝送時間内でより多くの端末を収容するために，周波数領域拡散を用いたGF-NOMAを検討した．直交周波数分割多重（OFDM）に基づく伝送モデルを想定し，単一のOFDMシンボルを用いて高精度なアクティブ端末および伝播路利得の推定を行う効率的な受信機を提案した．提案受信機では，遅延領域における伝播路のインパルス応答のスパース性を活用し，共分散推定で用いられるスパース再構成手法を拡張することで，無線伝播路の事前情報およびハイパーパラメータを必要とせずに高精度な推定を可能としている．数値結果より，圧縮センシングに用いられる既存のアルゴリズムと比較して，提案手法が優れた推定精度を達成できることを明らかにした．

最後に発展的検討として，第５章では，時間・周波数領域拡散を用いたGF-NOMAを想定し，近似メッセージ伝播法の性能解析に用いられる相転移解析に基づいたGF-NOMAの設計手法を提案した．時間・周波数資源を効率的に活用するため，パイロット信号とデータ信号のそれぞれに対して異なる拡散を施した伝送方法を用いる．また，遅延領域における伝播路のインパルス応答のスパース性を活用した信号モデルを導入することで，一般化MMV-AMPアルゴリズムを用いてアクティブ端末およびインパルス応答を効率的に推定できる．そして，推定したインパルス応答から補完して得られるデータ部に対応する伝播路利得を用いて，確率伝播法に基づいた低計算量なデータ推定が可能となる．計算機シミュレーションにより，１ミリ秒という低遅延制約下において，従来のOFDMに基づいたGF-NOMAの伝送法よりも優れた精度でアクティブ端末および伝播路利得を推定できることを示した．

これらの提案を通し，適切に設計されたGF-NOMAを用いることで多数接続・低遅延の両者を同時に満足し得ることを示した．本論文で得られた成果を活用することで，多数の無線端末が同時に低遅延で通信を行うことのできる無線通信技術の実現が期待できる．

Abstract

In this dissertation, we consider a grant-free non-orthogonal multiple access (GF-NOMA) system based on compressed sensing (CS), aiming to simultaneously satisfy massive connectivity and low latency. In GF-NOMA systems, each active user transmits data immediately without waiting for the grant, whereas the base station (BS) needs to perform active user detection and channel estimation to estimate the transmitted data correctly. We thus propose several GF-NOMA schemes to overcome the fundamental issues.

Firstly, we propose GF-NOMA schemes based on approximate message passing (AMP), which employs the transmission spreading over the time domain. We also design the threshold of active user detection (AUD) based on its performance analysis. It is shown that the proposed schemes can outperform or be comparable to the conventional one while lowering the computational complexity. After that, we consider the receivers for more realistic scenarios. Specifically, we propose the AUD based on the coordinate descent (CD) method taking into account carrier frequency offsets and the scheme to perform AUD and channel estimation (CE) based on AMP, which does not need prior knowledge of the channels. Numerical results show that both proposed methods attain comparable estimation accuracy to conventional ones.

To further reduce access latency, we introduce GF-NOMA schemes with spreading over the frequency domain and propose a hyperparameter-free receiver that exploits the CD method utilizing the channel sparsity in the delay domain while requiring no pre-tuning of parameters. It is revealed that the proposed receiver can perform AUD and CE more accurately than the conventional CS algorithms. Finally, we propose a feasible design of GF-NOMA that makes full use of time and frequency domains to accommodate more users. The proposed design is based on a tailored signal model and theoretical analyses of the family of AMP algorithms. We demonstrate that the proposed GF-NOMA can accommodate many users compared to the conventional scheme under the strict latency requirement.

Through these proposals, we address the possibility of realizing low latency communications by massive numbers of users, namely massive low-latency communications.

Table of Contents

List of Figures

List of Tables

List of Algorithms

List of Acronyms

3GPP third generation partnership project.

5G fifth generation.

5G NR 5G generation new radio.

AER activity error rate.

AMP approximate message passing.

AUD active user detection.

AWGN additive white Gaussian noise.

BCD block coordinate descent.

BER bit error rate.

BiGaBP bilinear Gaussian belief propagation.

BiGAMP bilinear generalized approximate message passing.

Boosted AMP boosted approximate message passing.

BP belief propagation.

BS base station.

BSASP block sparsity adaptive subspace pursuit.

CAMP complex approximate message passing.

CD coordinate descent.

CDMA code division multiple access.

CE channel estimation.

CFO carrier frequency offset.

CFR channel frequency response.

CIR channel impulse response.

CP cyclic prefix.

CRDSA contention resolution diversity slotted ALOHA.

CS compressed sensing.

DCS distributed compressed sensing.

DFT discrete Fourier transform.

EM expectation maximization.

EM-MMV-AMP expectation-maximization-based MMV-AMP.

FA false alarm.

FDMA frequency division multiple access.

FER frame error rate.

GaBP Gaussian belief propagation.

GF-NOMA grant-free non-orthogonal multiple access.

GMMV-AMP generalized MMV-AMP.

HyGAMP hybrid generalized approximate message passing.

i.i.d. independently and identically distributed.

IoT Internet-of-Things.

JACDE joint activity, channel, and data estimation.

JACE joint activity and channel estimation.

LLR log-likelihood ratio.

LS least squares.

LSF large-scale fading.

LTE long-term evolution.

M2M machine-to-machine.

MD miss detection.

MIMO multiple-input-multiple-output.

ML maximum likelihood.

MMSE minimum mean-squared error.

mMTC massive machine-type communications.

MMV-AMP multiple measurement vector approximate message passing.

MMVR multiple measurement vector reconstruction.

MPC multi-path component.

MRAS multi-rank aware sparse.

MSE mean-squared error.

MUD multiuser detection.

NB-IoT narrowband IoT.

NMSE normalized mean-squared error.

NNLS non-negative least square.

NOMA non-orthogonal multiple access.

OFDM orthogonal frequency division multiplexing.

OFDMA orthogonal frequency-division multiple access.

OMA orthogonal multiple access.

PDF probability density function.

PSK phase-shift keying.

QPSK quadrature phase-shift keying.

RB resource block.

ReMBo reduce MMV and boost.

ROC receiver operating characteristic.

SCMA sparse code multiple access.

SER symbol error rate.

SIC successive interference cancellation.

SMV single measurement vector.

SMVR single measurement vector recovery.

SNR signal-to-noise ratio.

TDMA time division multiple access.

TIN treating interference as noise.

ULA uniform linear array.

URLLC ultra reliable low latency communications.

Notation

$\mathbb{C}$	the field of complex numbers
$\mathbb{D}_+$	the domain of non-negative diagonal matrices
$\mathbb{R}$	the field of real numbers
$\mathbb{R}_+$	the field of non-negative real numbers
$\mathbb{E}[\cdot]$	the expectation operation
$\mathbb{E}_a[\cdot]$	the expectation operation over a
$\mathbb{E}[\cdot\|\cdot]$	the conditional expectation
$(\cdot)^*$	the conjugate of a complex value
$(\cdot)^{\mathrm{H}}$	the conjugate transpose of a vector/matrix
$(\cdot)^{\mathrm{T}}$	the transpose of a vector/matrix
$\mathbf{0}_N$	the N-dimensional all zeros vector
$\mathbf{1}_N$	the N-dimensional all ones vector
$\mathbf{I}_N$	the N-order identity matrix
$\mathbf{O}_{N,M}$	the $N \times M$ zero matrix
$\mathbf{X}^{-1}$	the inverse of square matrix $\mathbf{X}$ (if it exists)
$\| \mathbf{x} \|_p$	the ℓ_p-norm of vector $\mathbf{x}$
$\| \mathbf{X} \|_{\mathrm{F}}$	the Frobenius norm of matrix $\mathbf{X}$
$\det(\mathbf{X})$	the determinant of square matrix $\mathbf{X}$

$\mathrm{Tr}(\mathbf{X})$	the trace of square matrix $\mathbf{X}$
$\mathrm{diag}(\mathbf{x})$	the diagonal matrix with the elements of vector $\mathbf{x}$
$\mathrm{vec}(\mathbf{X})$	the row-major vectorization of matrix $\mathbf{X}$
$\lvert\mathcal{A}\rvert$	the cardinality of set $\mathcal{A}$
$\mathcal{A} \setminus \mathcal{B}$	the set whose elements are in set $\mathcal{A}$ but not in set $\mathcal{B}$
$\mathcal{CN}(\mu, \sigma^2)$	the circularly-symmetric complex-valued Gaussian distribution with the mean μ and the variance σ^2
$\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	the multi-dimensional complex-valued Gaussian distribution with the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$
$\mathcal{O}(\cdot)$	the big O notation, which is also known as Bachmann-Landau notation
$\lceil\cdot\rceil$	the ceiling function
$\otimes$	Kronecker product
$\log_a(x)$	the logarithm of x to base b
$\ln(x)$	the natural logarithm of x

Chapter 1

Introduction

This chapter first introduces the research background and then summarizes the development of multiple access schemes, showing the motivation to discuss grant-free multiple access schemes. After reviewing the related works, we state research motivations and contributions. Finally, we provide the organization of this dissertation.

1.1 Requirements for Future Multiple Access

Wireless communications technology becomes one of the essential infrastructures in our lives and plays the role of connecting people worldwide. In recent years, Internet-of-Things (IoT) has gained considerable attention because of its promising effects on our daily lives and industries [1] and plays the role of providing Internet access for a large number of devices pervasively. Accordingly, it has been predicted that together with the noticeable growth in machine-to-machine (M2M) connections, the number of mobile devices will reach 13.1 billion by 2023 [2], implying the coming of a new era that massive numbers of devices communicate for different uses. In this context, the third generation partnership project (3GPP) has selected massive machine-type communications (mMTC) as one of the major three use cases of the fifth generation (5G) wireless networks.

The traffic of IoT such as mMTC would be sporadic because only a small fraction of massive users will be active to transmit their sensing data, control data, and so on in each coherence time [3]. In these scenarios, extremely high energy efficiency is often required rather than data reliability and latency constraints. Meanwhile, some IoT applications, *e.g.*, automated driving and factory automation, impose stringent reliability and latency requirements to be demanded in ultra reliable low latency communications (URLLC). For instance, in factory automation applications, the cycle

time requirements for motion control and machine control vary from less than 0.5 msec to 10 msec [4], leading to the necessity of reducing the latency to be in the order of milliseconds.

From these backgrounds, future wireless communications systems will need to satisfy the heterogeneous requirements, especially the combination of mMTC and URLLC, namely massive URLLC [5–7]. In other words, they will be required to accommodate massive users over limited radio resources while coping with the sporadic IoT traffic and the strict latency requirements. In light of the above, let us summarize two main requirements for future multiple access as follows:

- Massive connectivity: Massive number of users should be accommodated.
- Low latency: Strict latency requirements need to be satisfied.

1.2 Conventional Random Access Schemes and their Limitations

To accommodate multiple users, multiple access is essential and has hence been evolved together with cellular systems. Previous and current cellular systems have employed a variety of orthogonal multiple access (OMA) techniques, such as frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA), and orthogonal frequency-division multiple access (OFDMA) [8–10]. The above OMA technologies exclusively assign radio resources to all users, which keeps orthogonality among all users and guarantees interference-free communication. However, such a resource allocation induces a considerable overhead when the number of users becomes massive [11]. To this end, *random access* approaches are attractive to accommodate large numbers of users.

Random access for grant-based transmissions

The cellular networks such as long-term evolution (LTE)-Advanced have employed the random access depicted in Fig. 1.1 for *grant-based* transmissions. As shown in Fig. 1.1, the random access in LTE-Advanced is performed to establish the connection between users and the base station (BS) ahead of uplink data transmission, whose procedure is based on the following four steps [12]: 1) each active user randomly picks a preamble from a predefined set of orthogonal preamble sequences and sends the preamble to the BS, playing the role of an access request; 2) Once the preamble is detected at the

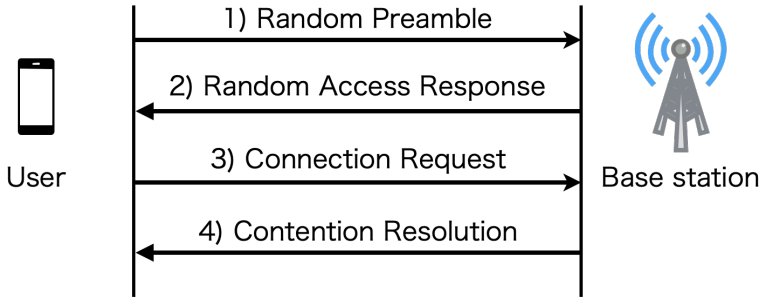

Fig. 1.1 An illustration of random access procedure in LTE-Advanced.

BS, a response corresponding the preamble is sent from the BS, which is a grant for transmitting in the next step; 3) The active user that has received a response sends a connection request to demand radio resources for data transmission; 4) If two or more users selected the same preamble in the first step, the BS detects a collision and then sends a contention resolution message as a final step to indicate whether or not the random access request from the device is granted. In addition to the above scheme, different variations of the random access schemes with advanced contention-resolution strategies have been recently investigated [13, 14].

However, as the number of collisions explosively grows with the number of users, still many users cannot access the network even if these schemes are utilized in the massive IoT scenarios [15]. Furthermore, the granting procedure results in a latency of around 9.5 msec ahead of the data transmission, which is far from the strict latency requirements of some IoT use cases [16]. Hence, instead of grant-based approaches, an alternative solution would be desired in the future IoT systems.

Conventional grant-free random access

To reduce access latency, *grant-free* random access has been extensively investigated [5, 6, 15]. This scheme employs the grant-free transmission that omits the granting process shown in Fig. 1.1, significantly reducing the signaling overhead ahead of data transmission. Hence, academia and industry have studied several grant-free approaches.

As one of the grant-free approaches, the 3GPP new radio Release 15 has supported the *K-repetition* scheme, which integrates the grant-free transmission with hybrid automatic repeat request [17]. In this scheme, each active user transmits the predefined number of consecutive replicas of the same packet without waiting for feedback from the BS. With a wide subcarrier spacing, the user can retransmit more replicas within the required latency, achieving the high-reliability levels. However, it was pointed out that the K-repetition scheme still has difficulty in supporting many users

with strict reliability and latency requirements since it suffers from a frequent collision between users that comes from the increase in the average number of packets generated per second [18]. In addition, this scheme is undesirable from a resource utilization efficiency point of view as it has to use a wider subcarrier spacing to satisfy the stringent latency requirements.

On the other hand, ALOHA-based protocols have been investigated for decades. In the 1970s, *pure ALOHA*, which considers a collision channel model with a single receiver and multiple transmitters and focuses on the contention resolution, was originated [19]. In this protocol, users immediately transmit their packets as soon as they are generated, while the receiver can demodulate the packets only if there is no collision. Hence, the maximum achievable throughput of pure ALOHA is limited to $1/2e \approx 0.18$, implying that only 18% of transmitted packets can be retrieved at the receiver. To alleviate the collisions, *slotted ALOHA* has been proposed [20], in which time slots are introduced, and each user transmits its packet within a time slot. The slot structure in slotted ALOHA can suppress partial collisions and then yields twice as high throughput as pure ALOHA. However, the probability of collisions is still high owing to incoordination among users, and then the system based on slotted ALOHA is likely to lack stability. Thus, different variants of ALOHA have been designed for better performance.

Specifically, in the 2000s, a paradigm shift in ALOHA-based protocols has arisen, following by the proposal of contention resolution diversity slotted ALOHA (CRDSA) [21]. CRDSA integrates inter-slot *successive interference cancellation (SIC)* and multiple transmission of packets with slotted ALOHA to retrieve more packets, significantly improving the achievable throughput. As an enhanced version of CRDSA, the schemes using *graph-based* design, coded slotted ALOHA, have been proposed [22, 23], which take notice of the fact that the resolution process of inter-slot SIC can be represented by a bipartite graph regularly utilized in coding theory. Thanks to this feature, coded slotted ALOHA can be well-designed by optimizing the probability mass function of the number of retransmissions, namely degree distribution, inspired by *density evolution* [24]. In addition, frameless ALOHA has been proposed [25] to further improve the performance when the number of active users fluctuate.

Although these approaches provide remarkably superior performance, the detection algorithm based on SIC requires a large memory to store all oversampled signals over multiple observation slots and results in extensive computational delays owing to iterative processing. These requirements are not preferable to IoT systems in practice.

1.3 New Paradigms to Accommodate Massive Users and Focus of This Dissertation

As mentioned in Section 1.2, conventional random access schemes have difficulty in accommodating a large number of users. Moreover, although multiple access channels, where multiple transmitters communicate with a common receiver, have been well-investigated from an information-theory perspective [26–28], these works assume the number of transmitters is fixed while the coding blocklength is allowed to be infinite. In addition, the conventional approaches have not taken into account the sporadic traffic in the IoT network. Thus, we need an alternative approach to model the random access problem for future IoT systems with short packet communications.

To address the issues above, two new communication paradigms have been developed, where the number of users grows with the coding blocklength unboundedly and the users are randomly activated: 1) *many-access channel* [29, 30] and 2) *unsourced random access* [31]. Notice that multiple access based on the model of [30] is often referred to as grant-free random access or *grant-free non-orthogonal multiple access (GF-NOMA)* in the related literature, *e.g.*, [6, 32].

In [30], the authors have characterized the channel capacity in the many-access channel under Gaussian noise, where the system assigns an *individual* codebook to each user. It is also shown that a two-phase coding strategy, which contains a pilot transmission phase for identifying active users and a subsequent data transmission phase, can achieve the capacity. This result corroborates the validity of a two-phase grant-free random access scheme in [15], where each active user transmits its pilot sequence in the first phase and immediately sends data in the second phase, as shown in Fig. 1.2. It is worth noting that the successive decoding strategy such as SIC, which can achieve the sum capacity in the conventional multiple access channels, is not applicable owing to the arbitrary large interference from the other users. Furthermore, the authors of [33] have extended the framework in [30] to multiple-input-multiple-output (MIMO) massive multiple access channels with non-symmetric rate. Theoretical analysis of the individual rate reveals that the number of antennas at the receiver asymptotically yields the degree of freedom gain, indicating no benefit of having multiple antennas at the transmitters.

On the other hand, the author of [31] has investigated the tradeoff between the energy-per-bit and the number of active users under the Gaussian multiple-access channel, where all users share a *common* codebook and each user sends information bits within a finite coding blocklength. The (massive) unsourced random access

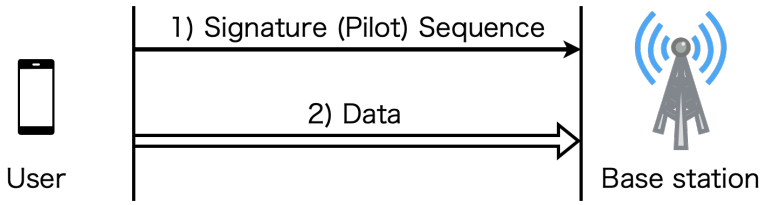

Fig. 1.2 An illustration of two-phase grant-free random access protocol.

framework reveals that the gap between the random coding achievability bound and existing schemes, *e.g.*, treating interference as noise (TIN), ALOHA, slotted ALOHA, and coded ALOHA, significantly increases with the number of active users, implying the necessity of the other promising solutions. Interestingly, this result implies that the SIC-based approaches are inefficient in massive IoT scenarios, the same as [30]. This fact simultaneously corroborates code-domain non-orthogonal multiple access (NOMA) [34, 35] is more suitable for such scenarios than power-domain NOMA [36, 37]. As a further advance, the authors of [38] and [39] have recently extended the framework of [31] into the quasi-fading channel and the block-fading MIMO channel, respectively.

Given the above, multiple access schemes based on the model of many-access channel or unsourced random access have been actively studied. Although unsourced random access is ongoing research topics and some sophisticate approaches have recently been proposed [40–44], they are likely to be only applicable to very low-rate mMTC scenarios, *e.g.*, 100 bits per user. Such a data rate dissatisfies the requirements for motion control applications, which vary from 20 byte to 50 byte. Therefore, we focus on GF-NOMA schemes to satisfy the requirements described in Section 1.1, namely massive connectivity and low latency.

1.4 Fundamental Challenges for GF-NOMA

GF-NOMA has the potential to accommodate many users efficiently but it has to cope with three fundamental challenges because of the grant-free fashion:

- Active user detection (AUD): The receiver has to identify active users because it does not know them in advance.
- Channel estimation (CE): The receiver requires the accurate channel state information to demodulate transmitted signals.
- Multiuser detection (MUD): The receiver needs to estimate transmitted signals from overlapped signals.

Thus, the promising schemes to solve these inherent problems would be required to realize massive grant-free access.

We briefly review the problems from a mathematical point of view. Let us consider a GF-NOMA system, where a BS equipped with M antennas serves K single-antenna users, out of which $K_{\mathrm{a}} \leq K$ users are active. An individual spreading sequence with length $L < K$, $\mathbf{a}_k \in \mathbb{C}^{L\times 1}$, is assigned to each user for uplink transmission. The channel between the BS and user k is denoted as $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$. Let $\mathcal{A} \subset \{1, 2, \ldots, K\}$ and s_k denote the set of active users' indices and the transmitted symbol of user k. Then, the received signal at the BS $\mathbf{Y} \in \mathbb{C}^{L\times M}$ is represented by

$$\begin{aligned}\mathbf{Y} &= \sum_{k\in\mathcal{A}} \mathbf{a}_k \underbrace{\mathbf{h}_k^{\mathrm{T}} s_k}_{\mathbf{x}_k^{\mathrm{T}}} + \mathbf{Z} \\ &= \mathbf{A}\mathbf{X} + \mathbf{Z} \end{aligned} \tag{1.1}$$

where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_K] \in \mathbb{C}^{L\times K}$ and $\mathbf{X} = [\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{K\times M}$ with $\mathbf{x}_k = s_k\mathbf{h}_k$, and $\mathbf{Z} \in \mathbb{C}^{L\times M}$ is a noise. As K_{a} is much smaller than K due to the sporadic traffic in massive IoT scenarios, $\mathbf{X}$ in (1.1) is a row sparse matrix. Thus, the estimation problem of $\mathbf{X}$ can be addressed using sparse recovery techniques such as *compressed sensing (CS)* [45].

The goals of AUD and CE are respectively to estimate $\mathcal{A}$ and $\mathbf{h}_k$ for $k \in \mathcal{A}$, where we have

$$s_k = \begin{cases} 1, & k \in \mathcal{A} \\ 0, & \text{Otherwise.} \end{cases} \tag{1.2}$$

This implies that each active user transmits the spreading sequence and then the BS uses it to identify the users. Notice that many researchers have addressed to perform AUD and CE jointly, namely joint activity and channel estimation (JACE).

On the other hand, the MUD problem is to estimate the data symbols transmitted by active users from the received signal using the estimates of $\mathcal{A}$ and $\mathbf{h}_k$. For the simplicity, we consider the scenario that each active user transmits the data symbol s_k multiplied by its own sequence $\mathbf{a}_k$ to the BS[1]. Here, we introduce $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{M\times K}$ and $\mathbf{s} = [s_1, \ldots, s_K]^{\mathrm{T}} \in \{\{0\} \cup \mathcal{X}\}^{K\times 1}$ with the set of data symbols $\mathcal{X}$. Then, (1.1) can be expressed by

$$\mathbf{Y} = \mathbf{A}\mathrm{diag}(\mathbf{s})\mathbf{H}^{\mathrm{T}} + \mathbf{Z}. \tag{1.3}$$

[1]Some existing approaches for MUD [46–48] do not need the multiplication of s_k and $\mathbf{a}_k$.

Moreover, applying the column-major vectorization to (1.3), we obtain the following relation:

$$\begin{aligned}\mathbf{y} &= (\mathbf{H} \otimes \mathbf{A})\mathbf{s}_{\mathrm{d}} + \mathbf{z} \\ &= \mathbf{H}_{\mathrm{eq}}\mathbf{s}_{\mathrm{d}} + \mathbf{z}, \end{aligned} \tag{1.4}$$

where $\mathbf{y} \in \mathbb{C}^{LM\times 1}$, $\mathbf{s}_{\mathrm{d}} \in \mathbb{C}^{K^2\times 1}$, and $\mathbf{z} \in \mathbb{C}^{LM\times 1}$ denote the vectors that stack the columns of $\mathbf{Y}$, diag($\mathbf{s}$), and $\mathbf{Z}$, respectively.

It is obvious from (1.4) that the MUD problem can be solved as the data detection in a multi-user MIMO system with the equivalent channel matrix $\mathbf{H}_{\mathrm{eq}} = (\mathbf{H} \otimes \mathbf{A}) \in \mathbb{C}^{LM\times K^2}$. Furthermore, the performance of MUD is significantly influenced by the estimation accuracy of $\mathcal{A}$ and $\mathbf{h}_k$. Therefore, many existing studies [49–64] have often tackled AUD (and CE). On the other hand, some existing works [46–48, 65–69] have investigated MUD together with AUD and CE, namely joint activity, channel, and data estimation (JACDE), to enhance the data estimation accuracy by exploiting the sparsity.

1.5 Related Work

In this section, we briefly review several conventional approaches tackling fundamental challenges in GF-NOMA systems.

1.5.1 AUD/JACE

AUD

For a narrowband GF-NOMA system, promising approaches for AUD have been proposed [49–52]. In [49], the authors have considered the AUD based on an approximate message passing (AMP) mechanism for the both the case where the case where the BS knows the exact large-scale fading (LSF)[2] coefficients as well as the case where only the statistics of them, *i.e.*, probability density function (PDF), is available. However, the full knowledge of the a priori distribution of the channels might not be available at the BS in practice.

In contrast to [49], in [50–52], the AUD problem has been formulated to the maximum likelihood (ML) estimation one and solved by the well-known coordinate

[2]In this thesis, we consider that the large-scale fading includes path-loss and shadowing, like the related literature. Notice that the path-loss and shadowing components are often distinguished [8].

descent (CD) algorithm, which only requires the sample covariance of the observations at the BS. Although this approach achieves superior performance without prior knowledge of the channels, its computational complexity grows quadratically with the pilot sequence length, inducing a prohibitive one in massive grant-free access.

As a further advance, the AUD in a frequency asynchronous GF-NOMA system has been addressed [53]. The proposed method in [53] uses the algorithm exploiting the Taylor expansion and the block coordinate descent (BCD) method but needs to know the maximum number of active users in advance, which the BS is unable to obtain in practical systems due to the nature of GF-NOMA.

JACE

As with AUD, the JACE schemes for a narrowband GF-NOMA system have been actively investigated. For instance, the authors of [54] have proposed the AUD and CE based on message passing mechanisms such as multiple measurement vector approximate message passing (MMV-AMP). Whereas the approach can perform AUD and CE efficiently, MMV-AMP requires the LSF coefficients of all users in advance. On the other hand, the authors of [55] have proposed the JACE based on a convex optimization with a regularization taking into account the row sparsity, while the authors of [56] have formulated the JACE problem to a low-rank sparse matrix recovery one, which is solved by a Riemannian optimization technique. However, the former needs an exhaustive search of regularization parameters for accurate estimation, and the latter only works under the assumption that the BS is equipped with antennas more than the number of active users. As a further advance, the JACE schemes for a time-asynchronous narrowband GF-NOMA system have recently investigated [57–60].

The aforementioned works utilize the non-orthogonal sequences spread over the *time domain*, and a BS estimates the active users from the overlapped measurements via sparse-recovery techniques. This transmission scheme has difficulty meeting the latency requirements unless the system bandwidth is widened since it requires a sufficiently long sequence in the time domain to accommodate massive users while keeping low AUD error. However, as the existing works [49–60] are designed for a narrowband system under frequency-flat fading channels, they are inapplicable for a wideband system under frequency-selective fading channels.

To this end, GF-NOMA systems with multicarrier transmission such as orthogonal frequency division multiplexing (OFDM) have been investigated in the related literature, *e.g.*, [61–64]. In [61], CS-based random access with multicarrier transmission and its associated pilot design were investigated. Nevertheless, the BS is assumed to be

able to utilize the number of active users and the maximum delay spread among all users. This assumption would be impractical owing to sporadic traffic and mobility of uplink users. Moreover, the authors of [62] have proposed the JACE based on multi-rank aware sparse (MRAS) recovery, which uses the inherent sparsity and low-rank structure of the millimeter-wave/Terahertz channels in the delay-angular domain, for millimeter-wave/Terahertz wideband GF-NOMA systems. Like [56], this MRAS recovery necessitates large numbers of antennas and the knowledge of the maximum number of propagation paths to exploit the low-rank structure. Thus, this scheme relies on a massive MIMO architecture at the BS, which is undesirable in terms of hardware cost. Furthermore, in [61] and [62], only one (OFDM) symbol is used to perform JACE, and then the number of users that the systems can accommodate is limited.

To alleviate the issue, the authors of [63] and [64] have proposed an alternate transmission approach using the generalized MMV-AMP (GMMV-AMP)-based algorithms, which exploit the inherent channel sparsity in both the spatial and angular domains. To further improve transmission performance, a pilot design based on a distributed compressed sensing (DCS) theory has been proposed. However, the basic GMMV-AMP algorithm in [63] does not take full advantage of the row sparsity that yields to performance gain. Furthermore, to perform accurate estimation in supporting massive users, the scheme in [64] requires not only a large number of antennas at the BS but also dozens of OFDM symbols. Especially, the use of dozens of OFDM symbols has difficulty in meeting the stringent latency requirements, *e.g.*, 1 msec, unless the subcarrier spacing is widened.

1.5.2 JACDE

Some existing approaches have considered all of the three challenges described in Section 1.4. For instance, the receiver for the single-antenna BS case have been investigated in [65–69]. In [66, 68], the authors have proposed the framework to perform AUD, CE, and MUD efficiently in grant-free sparse code multiple access (SCMA) systems. In addition, AMP-based approaches have been proposed in [67, 69]. Especially, the approach of [69] incorporated a rotationally invariant Gaussian mixture model into the message passing algorithm so as to further improve estimation accuracy. As an alternative scheme, the authors of [65] have considered the framework of JACDE based on block sparsity adaptive subspace pursuit (BSASP), which converts the original problem into a single measurement vector recovery (SMVR) problem with a block-sparse vector and then solved it.

In contrast to [65–69], the receivers for the multiple-antenna BS case, in which the problem of JACDE is solved as a *bilinear inference* problem, have been proposed [46–48]. In [46], the authors have proposed two schemes to perform JACDE based on bilinear generalized approximate message passing (BiGAMP) [70, 71] exploiting random sparsity learning or structured sparsity learning. Also, the scheme proposed in [47] uses the message passing algorithm that combines BiGAMP and loopy belief propagation (BP) inspired by hybrid generalized approximate message passing (HyGAMP) [72]. For uplink grant-free transmission in cell-free massive MIMO systems, the authors of [48] have proposed the receiver incorporating activity-awareness into the bilinear Gaussian belief propagation (BiGaBP) algorithm [73].

However, these conventional schemes have the following challenges: 1) BSASP [65] requires high-complexity operation, such as the calculation of a pseudo-inverse matrix, and properly tuned thresholds for each signal-to-noise ratio (SNR); 2) the approach in [69] considers a trilinear signal model tailored to the single-antenna BS case, which cannot be applied to the multiple-antenna BS case; 3) the receivers in [46–48] rely on the benefit of massive MIMO to accurately solve the formulated bilinear problem, requiring the use of expensive hardware.

1.6 Outline of This Dissertation

As described in the previous sections, although some promising approaches have been proposed, there are still practical issues to be overcome. These issues are summarized as follows:

1. Many JACDE schemes utilize the algorithms based on bilinear inference to attain superior performance, but the BS has to be equipped with tens or hundreds of antennas. Besides, some greedy schemes, *e.g.*, BSASP [65], need to calculate a pseudo-inverse matrix, resulting in prohibitive computational complexity in massive IoT scenarios.

2. Whereas the optimization-based JACE approaches proposed in [55] and [56] function without the knowledge of LSF coefficients, each of them suffers from the necessity of an exhaustive search of regularization parameters and the constraint that the BS must be equipped with antennas more than the number of active users. Moreover, GMMV-AMP [63] does not take full advantage of the row sparsity.

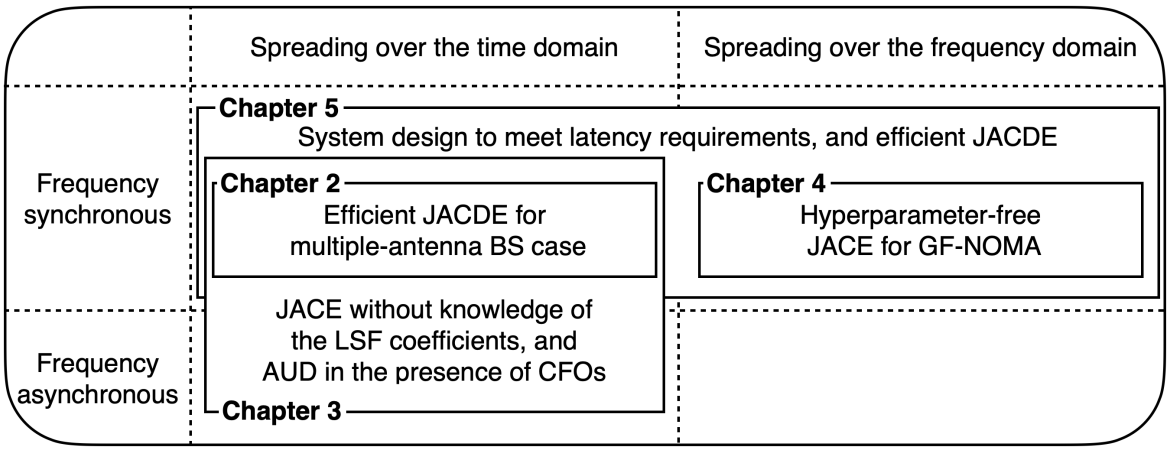

Fig. 1.3 Overview of this dissertation.

3. For frequency asynchronous scenarios, the challenges described in Section 1.4 have less been addressed. Although the AUD in the presence of carrier frequency offset (CFO)s was proposed in [53], it requires the maximum number of active users, which the BS cannot obtain in advance because of the grant-free nature.

4. To exploit the channel sparsity in the angular domain, conventional JACE schemes for OFDM-based GF-NOMA systems, *e.g.*, [62, 64], rely on the massive MIMO architecture that has to use extensive hardware. Also, the JACE scheme of [61] needs the number of active users and the maximum delay spread among all users.

5. Although GF-NOMA schemes based on multicarrier transmission have been studied [61, 62, 64], none of them have addressed the concrete system design to satisfy the latency requirements.

Therefore, this dissertation tackles the above issues by proposing several receivers and GF-NOMA schemes to meet massive connectivity and low latency, contributing to the establishment of massive low-latency communications.

The overview of this dissertation is shown in Fig. 1.3. As shown in Fig. 1.3, we consider three different spreading patterns throughout the dissertation. In Chapter 2, we first consider a narrowband GF-NOMA system with spreading over the time domain and propose two efficient receivers for the simple scenario. Chapter 3 then discusses the receivers for the above system under the more practical scenarios. Next, to further reduce access latency, Chapter 4 introduces an OFDM-based GF-NOMA system with spreading over the frequency domain and proposes its efficient receiver. After that, we propose the design of OFDM-based GF-NOMA systems that make full use of both the time and frequency domains to meet massive connectivity and low latency in

Chapter 5. Finally, Chapter 6 concludes the works in this dissertation. This dissertation is organized as follows.

Chapter 2: Efficient Receivers for GF-NOMA With Multiple-Antenna Base Station

In this chapter, we propose two efficient receivers for a narrowband GF-NOMA system with multiple-antenna BS to perform AUD, CE, and MUD. Each proposed receiver adopts *MMV-AMP* and *boosted approximate message passing (Boosted AMP)* respectively. We demonstrate that the symbol error rate (SER) performance of the receiver based on MMV-AMP approaches that of linear minimum mean-squared error (MMSE) with the perfect knowledge of active users by redesigning the threshold of AUD. Meanwhile, it is shown that the receiver based on Boosted AMP is comparable to the conventional scheme while significantly reducing the complexity.

Chapter 3: Receivers for GF-NOMA Against the Effects of Large-Scale Fading or CFOs

The previous chapter focuses on the case that the system can perform an ideal frequency synchronization and compensate for the effects of LSF of all users. On the other hand, this chapter respectively addresses two scenarios; 1) the BS does not know LSF coefficients of users and 2) CFOs exist. For the first scenario, we propose the algorithm named *expectation-maximization-based MMV-AMP (EM-MMV-AMP)*, which integrates the expectation maximization (EM) algorithm with the MMV-AMP algorithm to simultaneously estimate active users, channels, and LSF coefficients. For the second scenario, we formulate the problem of the AUD inspired by an idea of array-signal processing and then propose an approach based on the covariance-based CD method. It is shown through computer simulations that these proposed approaches can estimate active users and/or channel coefficients even when the systems cannot compensate for the effects of large-scale fading or CFOs.

Chapter 4: Hyperparameter-Free Receiver for GF-NOMA Using Frequency Domain

In two previous chapters, we have investigated GF-NOMA systems with spreading over the *time* domain. However, the transmission scheme employed in the systems requires a significant longer spreading sequence to efficiently support a large number of users. As a result, the systems have the difficulties in meeting the strict latency

requirements. To this end, we introduce a GF-NOMA system with the transmission spreading over the *frequency* domain, further reducing the latency. We also propose a *hyperparameter-free* receiver for the GF-NOMA systems, which is based on the iterative algorithm exploiting the sparsity of the channels in the *delay* domain. Numerical results demonstrate that the proposed scheme outperforms the conventional CS algorithms and that its performance is comparable to the state-of-the-art approach while avoiding the knowledge of the noise.

Chapter 5: Massive GF-NOMA Using Time and Frequency Domains

The above chapters mainly focus on GF-NOMA systems with spreading over time or frequency domain to accommodate many users. However, the performance of GF-NOMA is expected to be further improved if we can make GF-NOMA utilize both domains. In this chapter, we consider the design of GF-NOMA systems taking advantage of both the time and frequency domains so as to meet massive connectivity and low latency. We introduce a tailored signal model for properly enlarging the dimensionality of measurements and then propose the design based on an asymptotic analysis of a sparse recovery technique, namely *phase transition*. Moreover, the proposed GF-NOMA performs AUD and CE exploiting both the sparsity arising from sporadic traffic and channel sparsity in the delay domain, enabling efficient MUD. It is shown that GF-NOMA systems employing the proposed design can perform more accurate AUD and CE than the conventional OFDM-based scheme under the strict latency requirement. Furthermore, numerical results indicate that the proposed scheme has the potential to realize low latency communications while accommodating numbers of users. Finally, we discuss the advantageous region of each of the three spreading patterns considered in this dissertation.

Chapter 6: Conclusion and Future Works

In this chapter, we provide the conclusion and future works of this dissertation.

Chapter 2

Efficient Receivers for GF-NOMA With Multiple-Antenna Base Station

As introduced in Section 1.5.2, several JACDE schemes for GF-NOMA systems have been proposed [65–69, 46–48]. However, to attain superior performance, most existing schemes require a large number of antennas at the BS or calculation with prohibitive computational complexity.

In this chapter, to cope with the above issue, namely the first one described in Section 1.6, we propose two efficient receivers for a narrowband GF-NOMA system with a multiple-antenna BS. Herein, we consider the case that each active user autonomously transmits its own data by multiplying its unique but non-orthogonal spreading sequence. We first consider a receiver based on *MMV-AMP*, while designing the threshold for AUD based on the theoretical analysis. It is shown that the SER performance of the receiver using the designed threshold is comparable to that of grant-based NOMA in which the BS ideally knows active users in advance. Simultaneously, we propose a receiver based on *Boosted AMP*, which applies the conversion inspired by reduce MMV and boost (ReMBo) [74], to further reduce the complexity. This receiver converts the original multiple measurement vector reconstruction (MMVR) problem into a smaller-sized vector reconstruction problem and solves the one. Besides, it estimates channel coefficients and data symbols by a well-known linear MMSE detector after AUD via Boosted AMP. Numerical results reveal that the low-complexity receiver exhibits SER performance comparable to that of a conventional scheme.

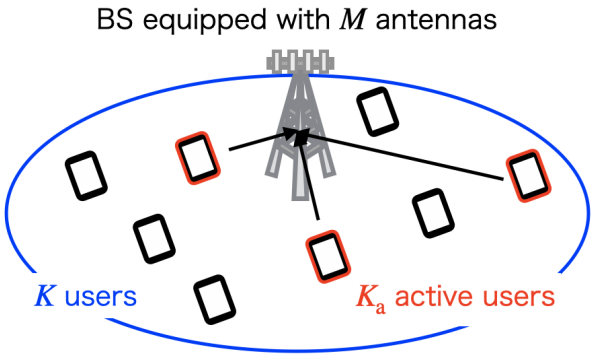

Fig. 2.1 Uplink GF-NOMA system.

2.1 System model

We consider a model comprising a network with K single-antenna users and a common BS equipping M antennas, as shown in Fig. 2.1. Each user has its unique but non-orthogonal spreading sequence of length $L < K$, exploiting the transmission scheme of code-domain NOMA. Throughout this chapter, we assume that all users are time and frequency synchronized, only $K_\mathrm{a} \leq K$ users are active within the transmission duration, and channels do not change in the duration. As shown in Fig. 2.2, every active user transmits the frame that consists of a pilot symbol and J data symbols multiplied by their own spreading sequence to the common BS, whereas the others do not transmit any symbols in the duration.

Let $x_{k,\mathrm{p}} \in \mathbb{C}$, $x_{k,\mathrm{d}}^{(j)} \in \mathbb{C}$, and $\mathbf{a}_k \in \mathbb{C}^{L\times 1}$ be the pilot symbol, the data symbol at j th time slot, and the unique spreading sequence for user k, respectively. Then, the received signals in the entire frame at BS, $\mathbf{Y} = [\mathbf{Y}_\mathrm{p}, \mathbf{Y}_\mathrm{d}^{(1)}, \ldots, \mathbf{Y}_\mathrm{d}^{(J)}] \in \mathbb{C}^{L\times M(J+1)}$, can be expressed as

$$
\begin{aligned}
\mathbf{Y} &= \sqrt{\xi}\sum_{k\in\mathcal{A}} \mathbf{a}_k \left[\underbrace{x_{k,\mathrm{p}}\mathbf{h}_k^\mathrm{T}}_{\mathbf{x}_{k,\mathrm{p}}}, \underbrace{x_{k,\mathrm{d}}^{(1)}\mathbf{h}_k^\mathrm{T}}_{\mathbf{x}_{k,\mathrm{d}}^{(1)}}, \ldots, \underbrace{x_{k,\mathrm{d}}^{(J)}\mathbf{h}_k^\mathrm{T}}_{\mathbf{x}_{k,\mathrm{d}}^{(J)}} \right] + \left[\mathbf{Z}_\mathrm{p}, \mathbf{Z}_\mathrm{d}^{(1)}, \ldots, \mathbf{Z}_\mathrm{d}^{(J)}\right] \\
&= \sqrt{\xi}\left[\mathbf{a}_1, \ldots, \mathbf{a}_K\right] \begin{bmatrix} \left[\mathbf{x}_{1,\mathrm{p}}, \mathbf{x}_{1,\mathrm{d}}^{(1)}, \ldots, \mathbf{x}_{1,\mathrm{d}}^{(J)}\right] \\ \vdots \\ \left[\mathbf{x}_{K,\mathrm{p}}, \mathbf{x}_{K,\mathrm{d}}^{(1)}, \ldots, \mathbf{x}_{K,\mathrm{d}}^{(J)}\right] \end{bmatrix} + \mathbf{Z} \\
&= \sqrt{\xi}\mathbf{A}\left[\mathbf{X}_\mathrm{p}, \mathbf{X}_\mathrm{d}^{(1)}, \ldots, \mathbf{X}_\mathrm{d}^{(J)}\right] + \mathbf{Z} \\
&= \sqrt{\xi}\mathbf{A}\mathbf{X} + \mathbf{Z}, \qquad (2.1)
\end{aligned}
$$

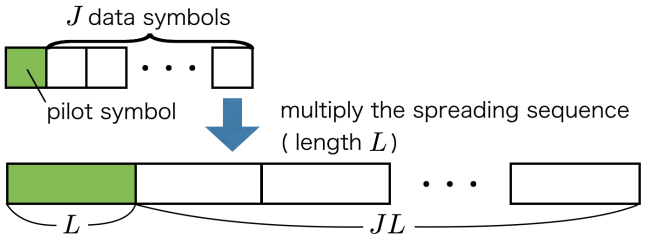

Fig. 2.2 Illustration of the frame structure.

where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_K] \in \mathbb{C}^{L\times K}$, $\mathbf{X} = [\mathbf{X}_\mathrm{p}, \mathbf{X}_\mathrm{d}^{(1)}, \ldots, \mathbf{X}_\mathrm{d}^{(J)}] \in \mathbb{C}^{L\times M(J+1)}$, and $\mathbf{Z} = [\mathbf{Z}_\mathrm{p}, \mathbf{Z}_\mathrm{d}^{(1)}, \ldots, \mathbf{Z}_\mathrm{d}^{(J)}] \in \mathbb{C}^{L\times M(J+1)}$. Notice that $\mathbf{Y}_\mathrm{p} = \sqrt{\xi}\mathbf{A}\mathbf{X}_\mathrm{p} + \mathbf{Z}_\mathrm{p}$ and $\mathbf{Y}_\mathrm{d}^{(j)} = \sqrt{\xi}\mathbf{A}\mathbf{X}_\mathrm{d}^{(j)} + \mathbf{Z}_\mathrm{d}^{(j)}$ represent the received signals at the pilot phase and the j th data phase, respectively. In addition, ξ, $\mathcal{A}$, and $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ denote the transmission power of each active user in each time slot, the set of indices of active users, and the channel vector of user k, respectively. The channel vectors are assumed to be independent from each other and are uncorrelated along the antennas, *i.e.*, $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}_M, \beta_k \mathbf{I}_M)$ with the LSF coefficient β_k. The matrices $\mathbf{X}_\mathrm{p} = [\mathbf{x}_{1,\mathrm{p}}^\mathrm{T}, \ldots, \mathbf{x}_{K,\mathrm{p}}^\mathrm{T}]^\mathrm{T} \in \mathbb{C}^{K\times M}$ and $\mathbf{X}_\mathrm{d}^{(j)} = [(\mathbf{x}_{1,\mathrm{d}}^{(j)})^\mathrm{T}, \ldots, (\mathbf{x}_{K,\mathrm{d}}^{(j)})^\mathrm{T}]^\mathrm{T} \in \mathbb{C}^{K\times M}$ are equivalent signals at the pilot phase and the j th data phase, where $\mathbf{x}_{k,\mathrm{p}} = x_{k,\mathrm{p}}\mathbf{h}_k^\mathrm{T}$ and $\mathbf{x}_{k,\mathrm{d}}^{(j)} = x_{k,\mathrm{d}}^{(j)}\mathbf{h}_k^\mathrm{T}$. Moreover, $\mathbf{Z}_\mathrm{p} \in \mathbb{C}^{L\times M}$ and $\mathbf{Z}_\mathrm{d}^{(j)} \in \mathbb{C}^{L\times M}$ for $j = 1, 2, \ldots, J$ are noise matrices in which all elements obey independently and identically distributed (i.i.d.) complex Gaussian distribution with zero mean and variance σ_n^2.

It is obvious from (2.1) that the equivalent signals $\mathbf{X}$ include both transmitted symbols and channel effects. Without loss of generality, for $k \in \mathcal{A}$, we can respectively set the pilot symbol $x_{k,\mathrm{p}}$ and data symbol $x_{k,\mathrm{d}}^{(j)}$ to one and the Q-ary phase-shift keying (PSK) symbol with normalized power, otherwise the symbols are supposed to be zero. In light of the above, the equivalent signals $\mathbf{X}$ can be regarded as the row-sparse matrix.

2.2 Proposed Receivers

2.2.1 Preliminaries

We first explain a problem formulation in the conventional approach [65] to clarify the difference between conventional and proposed receivers.

In [65], the original form (2.1) is converted into the SMVR problem with a single block sparse signal as follows:

$$\begin{aligned}\mathrm{vec}(\mathbf{Y}) &= \sqrt{\xi}\mathrm{vec}(\mathbf{A}\mathbf{X}) + \mathrm{vec}(\mathbf{Z}) \\ &= \sqrt{\xi}(\mathbf{A} \otimes \mathbf{I}_{M(J+1)})\mathrm{vec}(\mathbf{X}) + \mathrm{vec}(\mathbf{Z}),\end{aligned} \tag{2.2}$$

where $\mathrm{vec}(\cdot)$ denotes the row-major vectorization.[1] As the vector $\mathrm{vec}(\mathbf{X}) \in \mathbb{C}^{KM(J+1)\times 1}$ in (2.2) can be regarded as the K_a block sparse signal composed of K blocks with block size $M(J+1)$, the conventional receiver solves the problem (2.2) by BSASP.

To solve the SMVR problem in (2.2), the single measurement vector (SMV) version of MMV-AMP, such as AMP algorithm [76], can be applied. However, the performance of AMP significantly degrades since the equivalent measurement matrix $\mathbf{A} \otimes \mathbf{I}_{M(J+1)} \in \mathbb{C}^{LM(J+1)\times KM(J+1)}$ consists of multiple diagonal matrices and these elements are not independently and identically distributed [77].

On the other hand, our proposed receivers straightforwardly deal with the original problem (2.1). We assume that the rows of the matrix $\mathbf{X}$ in (2.1) follows the *Bernoulli Gaussian* distribution

$$p(\mathbf{x}_k) = (1-\epsilon)\delta_0(\mathbf{x}_k) + \epsilon p(\mathbf{h}), \quad {}^\forall k, \tag{2.3}$$

where $\epsilon \triangleq K_\mathrm{a}/K$, $\delta_0(\cdot)$ denotes Dirac's delta function, and $p(\mathbf{h})$ denotes the distribution of the rows corresponding to active users, such as $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}_{M(J+1)}, \beta_k \mathbf{I}_{M(J+1)})$. This assumption is based on the fact that the k ($\in \mathcal{A}$) th row vector of $\mathbf{X}$ consists of the products of the channel $\mathbf{h}_k$ and complex symbols, which become one or a certain PSK symbol whose energy is normalized. The equivalent measurement matrix $\mathbf{A}$ can satisfy the condition that the elements are independent identically distributed and thus enables the BS to use the low-complexity algorithm based on AMP to reconstruct the matrix $\mathbf{X}$. Therefore, the proposed receivers can conduct AUD, CE, and MUD with lower complexity than the conventional one. The following subsections explain the proposed receivers, which are based on MMV-AMP and Boosted AMP, under the above assumption.

[1] In general, the operator $\mathrm{vec}(\cdot)$ express the column-major vectorization, which stacks the columns into a long column vector [75].

2.2.2 Receiver Based on MMV-AMP

MMV-AMP algorithm

Our proposed receiver estimates not only channel coefficients but data symbols corresponding to active users simultaneously, unlike [54]. Here, we first explain the basics of the MMV-AMP algorithm to realize this estimation.

Initializing $\mathbf{R}^{(0)} = \mathbf{Y}$ and $\mathbf{X}^{(0)} = \mathbf{O}_{K\times M(J+1)}$, the operations of the MMV-AMP algorithm at each iteration can be expressed as

$$\mathbf{x}_k^{(t+1)} = \eta_{t,k}\left((\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}\right), \tag{2.4}$$

$$\mathbf{R}^{(t+1)} = \mathbf{Y} - \mathbf{A}\mathbf{X}^{(t+1)} + \frac{K}{L}\mathbf{R}^{(t)}\sum_{k=1}^{K}\frac{\eta'_{t,k}\left((\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}\right)}{K}, \tag{2.5}$$

where $t = 0, 1, \ldots$ is the index of the iteration, $\mathbf{X}^{(t)} = [\mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_K^{(t)}]^{\mathrm{T}} \in \mathbb{C}^{K\times M(J+1)}$ is the estimate of $\mathbf{X}$ at iteration t, and $\mathbf{R}^{(t)} = [\mathbf{r}_1^{(t)}, \ldots, \mathbf{r}_{M(J+1)}^{(t)}] \in \mathbb{C}^{L\times M(J+1)}$ denotes the corresponding residual. Moreover, $\eta_{t,k}(\cdot) : \mathbb{C}^{M(J+1)\times 1} \mapsto \mathbb{C}^{M(J+1)\times 1}$ is the denoiser, and $\eta'_{t,k}(\cdot) \in \mathbb{C}^{M(J+1)\times M(J+1)}$ is the first-order derivative of $\eta_{t,k}(\cdot)$.

MMSE denoiser

The design of $\eta_{t,k}(\cdot)$ significantly influences the estimation accuracy of the algorithm, and thus, we consider using the *MMSE denoiser* [54]. An MMSE denoiser is derived based on a prior model of $\mathbf{X}$ in (2.3) and then given by the conditional expectation $\mathbb{E}[\boldsymbol{X}_k|\bar{\boldsymbol{X}}_{t,k} = \bar{\mathbf{x}}_{t,k}]$. Here, $\bar{\boldsymbol{X}}_{t,k} = \boldsymbol{X}_k + \boldsymbol{\Sigma}_t^{\frac{1}{2}}\boldsymbol{V}_k$ are defined with $\boldsymbol{X}_k \in \mathbb{C}^{M(J+1)\times 1}$ following the distribution in (2.3), $\boldsymbol{V}_k \sim \mathcal{CN}(\mathbf{0}_{M(J+1)}, \mathbf{I}_{M(J+1)})$, and the positive definite matrix $\boldsymbol{\Sigma}_t$ [54].

For the sake of simplicity, the positive definite matrix is assumed to be a scaled identity matrix at each iteration.

$$\boldsymbol{\Sigma}_t = \tau_t^2\mathbf{I}_{M(J+1)}, \quad {}^{\forall}t \geq 0, \tag{2.6}$$

where we have

$$\tau_t^2 \triangleq \frac{\|\mathbf{R}^{(t)}\|_{\mathrm{F}}^2}{LM(J+1)}. \tag{2.7}$$

In addition, owing to (2.6), the signal $\bar{\mathbf{x}}_{t,k}$ can be modeled as

$$\bar{\mathbf{x}}_{t,k} = \mathbf{x}_k + \tau_t \mathbf{v}_k, \tag{2.8}$$

with $\mathbf{v}_k \sim \mathcal{CN}(\mathbf{0}_{M(J+1)}, \mathbf{I}_{M(J+1)})$. Hence, the MMSE denoiser can be expressed as

$$\begin{aligned}\eta_{t,k}(\bar{\mathbf{x}}_{t,k}) &= \mathbb{E}[\boldsymbol{X}_k | \bar{\boldsymbol{X}}_{t,k} = \bar{\mathbf{x}}_{t,k}] \\ &= \phi_{t,k} \frac{\beta_k}{\beta_k + \tau_t^2} \bar{\mathbf{x}}_{t,k}, \quad {}^{\forall} t, k,\end{aligned} \tag{2.9}$$

where

$$\phi_{t,k} = \frac{1}{1 + \frac{1-\epsilon}{\epsilon} \exp\left(-M(J+1)(\pi_{t,k} - \psi_{t,k})\right)}, \tag{2.10}$$

$$\pi_{t,k} = \frac{1}{M(J+1)} \left(\frac{1}{\tau_t^2} - \frac{1}{\beta_k + \tau_t^2} \right) \|\bar{\mathbf{x}}_{t,k}\|_2^2, \tag{2.11}$$

$$\psi_{t,k} = \ln\left(1 + \frac{\beta_k}{\tau_t^2}\right). \tag{2.12}$$

Furthermore, the random vector $\bar{\boldsymbol{X}}_{t,k}$ can be written as $\bar{\boldsymbol{X}}_{t,k} = \boldsymbol{X}_k + \tau_t \boldsymbol{V}_k$ because of the assumption of (2.6). Therefore, the first-order derivative of the MMSE denoiser $\eta_{t,k}(\cdot)$ is [78]:

$$\eta'_{t,k}(\bar{\mathbf{x}}_{t,k}) = \frac{1}{\tau_t^2} \mathrm{var}[\boldsymbol{X}_k | \bar{\boldsymbol{X}}_{t,k} = \bar{\mathbf{x}}_{t,k}], \tag{2.13}$$

where $\mathrm{var}[\boldsymbol{X}_k | \bar{\boldsymbol{X}}_{t,k} = \bar{\mathbf{x}}_{t,k}]$ denotes the conditional variance of $\boldsymbol{X}_k$. Based on [54], the conditional variance of $\boldsymbol{X}_k$ can be expressed as

$$\begin{aligned}&\mathrm{var}[\boldsymbol{X}_k | \bar{\boldsymbol{X}}_{t,k} = \bar{\mathbf{x}}_{t,k}] \\ &= \mathbb{E}_{\boldsymbol{X}_k | \bar{\boldsymbol{X}}_{t,k} = \bar{\mathbf{x}}_{t,k}} [(\eta_{t,k}(\bar{\mathbf{x}}_{t,k}) - \bar{\mathbf{x}}_{t,k})(\eta_{t,k}(\bar{\mathbf{x}}_{t,k}) - \bar{\mathbf{x}}_{t,k})^{\mathrm{H}}] \\ &= \frac{1}{\tau_t^2} \left\{ \frac{\phi_{t,k} \beta_k \tau_t^2}{\beta_k + \tau_t^2} \mathbf{I}_{M(J+1)} + \frac{\phi_{t,k}(1 - \phi_{t,k}) \beta_k^2}{(\beta_k + \tau_t^2)^2} \bar{\mathbf{x}}_{t,k} \bar{\mathbf{x}}_{t,k}^{\mathrm{H}} \right\}.\end{aligned} \tag{2.14}$$

Estimation of active users, channel coefficients, and data symbols

After the process of the MMV-AMP algorithm, the AUD is first performed based on the characteristics of the MMSE denoiser. The estimates of channel coefficients and data symbols are then obtained from the outputs of MMV-AMP, simultaneously.

The AUD exploits the active user detector that is designed based on the fact that $\phi_{t,k}$ given in (2.10) tends to 1 if $\pi_{t,k} > \psi_{t,k}$ and 0 if $\pi_{t,k} < \psi_{t,k}$ for large $M(J+1)$. Thus, the estimated set of $\mathcal{A}$ is determined as

$$\hat{\mathcal{A}} = \left\{ k \mid \|(\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}\|_2^2 > \theta_{t,k}, \quad k \in \{1, \ldots, K\} \right\}, \tag{2.15}$$

with the threshold $\theta_{t,k}$ given by

$$\theta_{t,k} = \left(\frac{1}{\tau_t^2} - \frac{1}{\beta_k + \tau_t^2} \right)^{-1} M(J+1) \ln \left(1 + \frac{\beta_k}{\tau_t^2} \right). \tag{2.16}$$

The derivation of $\theta_{t,k}$ will be described in Section 2.3.

After the AUD based on (2.15), the channel coefficients corresponding to user $k \in \hat{\mathcal{A}}$ are estimated as

$$\hat{\mathbf{h}}_k = [x_{k,1}^{(t)}, \ldots, x_{k,M}^{(t)}]^{\mathrm{T}}. \tag{2.17}$$

Finally, data symbol of user k at the j-th time slot is determined by the following criterion:

$$\hat{x}_{k,\mathrm{d}}^{(j)} = \underset{x \in \mathcal{X}}{\arg\min} \ \|x\hat{\mathbf{h}}_k - [x_{k,Mj+1}^{(t)}, \ldots, x_{k,Mj+M}^{(t)}]^{\mathrm{T}}\|_2, \quad j = 1, \ldots, J, \tag{2.18}$$

where $\mathcal{X}$ denotes the set of the modulated symbols, namely the PSK symbols.

2.2.3 Low-Complexity Receiver Based on Boosted AMP

Although the receiver based on MMV-AMP exhibits superior performance with low complexity, its complexity is still high when the number of measurements increases considerably. To cope with this impediment, we propose a receiver based on Boosted AMP. In a similar fashion to the two-stage approach [67], this proposed receiver executes the AUD using the Boosted AMP and joint CE and MUD using the MMSE detector. Thus, this receiver performs the following steps:

1. Convert the MMVR problem into the one whose dimension is smaller than M with the same sparsity pattern by multiplying the received signal $\mathbf{Y}$ by a random matrix $\mathbf{V}_i \in \mathbb{C}^{M(J+1) \times M_{\mathrm{B}}}$, *i.e.*, $\tilde{\mathbf{Y}}_i = \mathbf{Y}\mathbf{V}_i$, where $M_{\mathrm{B}} \leq M(J+1)$ and i denote the number of columns in the converted signal and the index of the iteration, respectively.

2. Solve the converted MMVR problem using the MMV-AMP algorithm. Starting with $\tilde{\mathbf{R}}^{(0)} = \tilde{\mathbf{Y}}_i \in \mathbb{C}^{L\times M_\mathrm{B}}$ and $\tilde{\mathbf{X}}^{(0)} = \mathbf{O}_{k\times M_\mathrm{B}} \in \mathbb{C}^{K\times M_\mathrm{B}}$, we compute the matrices $\tilde{\mathbf{R}}_i^{(t)} \in \mathbb{C}^{L\times M_\mathrm{B}}$ and $\tilde{\mathbf{X}}_i^{(t)} \in \mathbb{C}^{K\times M_\mathrm{B}}$ using (2.4) and (2.5), instead of $\mathbf{R}^{(t)}$ and $\mathbf{X}^{(t)}$ as defined in Section 2.2.2. For simplicity, we assume that the desired signals of the MMVR problem follow the Bernoulli Gaussian distribution.

3. After the iterations of the MMV-AMP algorithm, the log-likelihood ratio (LLR) L_k^i for AUD is calculated as

$$L_k^i = M_\mathrm{B} \ln\left(\frac{\tau_t^2}{\beta_k+\tau_t^2}\right) + \left(\frac{1}{\tau_t^2} - \frac{1}{\beta_k+\tau_t^2}\right)\left\|\tilde{\mathbf{x}}_{i,k}^{(t)}\right\|_2^2 \tag{2.19}$$

   where $\tilde{\mathbf{x}}_{i,k}^{(t)} = (\tilde{\mathbf{R}}_i^{(t)})^\mathrm{T}\mathbf{a}_k^* + \tilde{\mathbf{x}}_{i,k}^{(t)}$.

4. Iterate 1) to 3) until i reaches the maximum number of iterations T_rem.

5. Set the estimated set of $\mathcal{A}$ as

$$\hat{\mathcal{A}} := \left\{ k \;\middle|\; \sum_{i=1}^{T_\mathrm{rem}} L_k^i > 0,\ k \in \{1,\ldots,K\} \right\}. \tag{2.20}$$

6. Based on $\hat{\mathcal{A}}$ and $\mathbf{Y}$, joint CE and MUD are performed using the MMSE detector.

$$\hat{\mathbf{X}} = \left(\tilde{\mathbf{A}}^\mathrm{H}\tilde{\mathbf{A}} + \sigma_k^2\mathbf{I}_{|\hat{\mathcal{A}}|}\right)^{-1}\tilde{\mathbf{A}}^\mathrm{H}\mathbf{Y}, \tag{2.21}$$

   where the matrix $\tilde{\mathbf{A}}$ is the matrix comprised of the column vectors $\mathbf{a}_k$ with $k \in \hat{\mathcal{A}}$. Therefore, the estimated channels and data symbols can be obtained as

$$\hat{\mathbf{h}}_k = [\hat{x}_{k,1},\ldots,\hat{x}_{k,M}]^\mathrm{T}, \tag{2.22}$$

$$\hat{x}_{k,d}^{(j)} = \arg\min_{x\in\mathcal{X}} \left\| x\hat{\mathbf{h}}_k - [\hat{x}_{k,jM+1},\ldots,\hat{x}_{k,jM+M}]^\mathrm{T}\right\|_2, \tag{2.23}$$

   respectively.

The converted problem when $M_\mathrm{B} = 1$ is SMVR, which is a special case of MMVR, and can be solved via AMP [76], enabling to further reduce the computational complexity of AUD. However, the accuracy of AUD degrades compared to Boosted AMP with $M_\mathrm{B} > 1$, and thus, there is a tradeoff between its performance and complexity. The tradeoff will be discussed in Section 2.4.

Table 2.1 Computational complexities of the conventional and proposed receivers.

Receiver	Complexity
MMV-AMP	$\mathcal{O}(KLM(J+1))$
Boosted AMP	$\mathcal{O}(KLM_{\mathrm{B}}T_{\mathrm{rem}})$
BSASP [65]	$\mathcal{O}(KL(M(J+1))^2+(M(J+1)s_i)^3)$

2.2.4 Complexity Comparisons

In this subsection, we briefly review the complexity of each receiver. The operation in the MMV-AMP algorithm includes the matrix multiplication of (2.5), namely $\mathbf{A}\mathbf{X}^{(t+1)}$, dominating the computational complexity of the receiver based on MMV-AMP. As $\mathbf{A}\in\mathbb{C}^{L\times K}$ and $\mathbf{X}^{(t+1)}\in\mathbb{C}^{K\times M(J+1)}$, its complexity order is given by $\mathcal{O}(KLM(J+1))$.

In contrast, the complexity of the receiver based on Boosted AMP depends on the number of iterations to solve the converted MMVR problem by MMV-AMP. In other words, the computational complexity comes from the iteration of the matrix multiplication of the $L\times K$ and $K\times M_{\mathrm{B}}$ matrices. The overall complexity of the receiver based on Boosted AMP is hence given by $\mathcal{O}(KLM_{\mathrm{B}}T_{\mathrm{rem}})$. It is worth noting that the parameters M_{B} and T_{rem} are typically set to satisfy $M_{\mathrm{B}}T_{\mathrm{rem}}<M(J+1)$ to reduce complexity.

Finally, we describe the required complexity of the receiver based on BSASP. BSASP requires the number of complex multiplications in each iteration [65]: $5K+3KM(J+1)+3KL(M(J+1))^2+2LM(J+1)(M(J+1)s_i)^2+(M(J+1)s_i)^3$, where s_i is the number of estimated active users in the i-th iteration. Thus, we conclude that the main computational complexity of the BSASP is expressed as $\mathcal{O}(KL(M(J+1))^2+(M(J+1)s_i)^3)$, which grows exponentially as the number of measurements increase, *i.e.*, antennas and transmitted symbols.

In light of the above, the complexities of the conventional and proposed receivers are listed in Table 2.1. This table indicates that our proposed receivers can remarkably reduce the computational complexity compared with BSASP [65], and thus, they have the potential to apply the GF-NOMA systems with massive MIMO.

2.3 Performance Analyses and Threshold Design for AUD

The performance of our proposed receivers depends on the performance of AUD because the estimation of channels and data symbols is performed based on the results of AUD. In this section, we analyze the AUD performance of the proposed receivers in terms of miss detection (MD) probability, which is the probability that an active user is incorrectly detected as a non-active user and is defined as

$$P^{\mathrm{MD}}(M) \triangleq \mathbb{E}\left[\frac{1}{K_{\mathrm{a}}}|\mathcal{A} \setminus \hat{\mathcal{A}}|\right]. \tag{2.24}$$

Moreover, the threshold $\theta_{t,k}$ in (2.15) is most important to perform AUD accurately, and it affects the overall performance of GF-NOMA systems.

2.3.1 MMV-AMP

As both proposed receivers exploit the MMV-AMP algorithm, we first describe the performance analysis of the receiver based on MMV-AMP.

To analyze the performance of our proposed receiver, we utilize *state evolution* that is a well-known tool to provide the performance limit of the AMP algorithm assuming a *large system limit* [76, 79, 80]. Under this assumption, the number of users K, the number of active users K_{a}, and the length of spreading sequences L all go to infinity, while their ratios K/L and K_{a}/K converge to some fixed positive values $\omega \in (0, \infty)$ and $\epsilon \in (0, \infty)$, respectively. Given the fixed transmission power ξ, the theoretical value of τ_t^2 can be iteratively computed by [54]:

$$\tau_0^2 = \frac{\sigma_{\mathrm{n}}^2}{\xi} + \omega\epsilon\mathbb{E}_\beta[\beta], \tag{2.25}$$

$$\tau_{t+1}^2 = \frac{\sigma_{\mathrm{n}}^2}{\xi} + \omega\epsilon\mathbb{E}_\beta\left[\frac{\beta\tau_t^2}{\beta+\tau_t^2}\right] + \omega\mathbb{E}_\beta\left[\vartheta_{t,\beta}(\tau_t^2)\right], \quad t \geq 0, \tag{2.26}$$

where we have

$$\vartheta_{t,\beta}(\tau_t^2) = \frac{1}{M(J+1)}\mathbb{E}_{\bar{\boldsymbol{X}}_{t,\beta}}\left[\frac{\phi_{t,k}(1-\phi_{t,\beta})\beta^2}{(\beta+\tau_t^2)^2}\bar{\boldsymbol{X}}_{t,\beta}^{\mathrm{H}}\bar{\boldsymbol{X}}_{t,\beta}\right]. \tag{2.27}$$

Here, $\bar{\boldsymbol{X}}_{t,\beta}$ denotes the random vector obeying the distribution of the signal $\bar{\mathbf{x}}_{t,k}$ given in (2.8), and $\phi_{t,\beta}$ captures the distribution of $\phi_{t,k}$ given in (2.10).

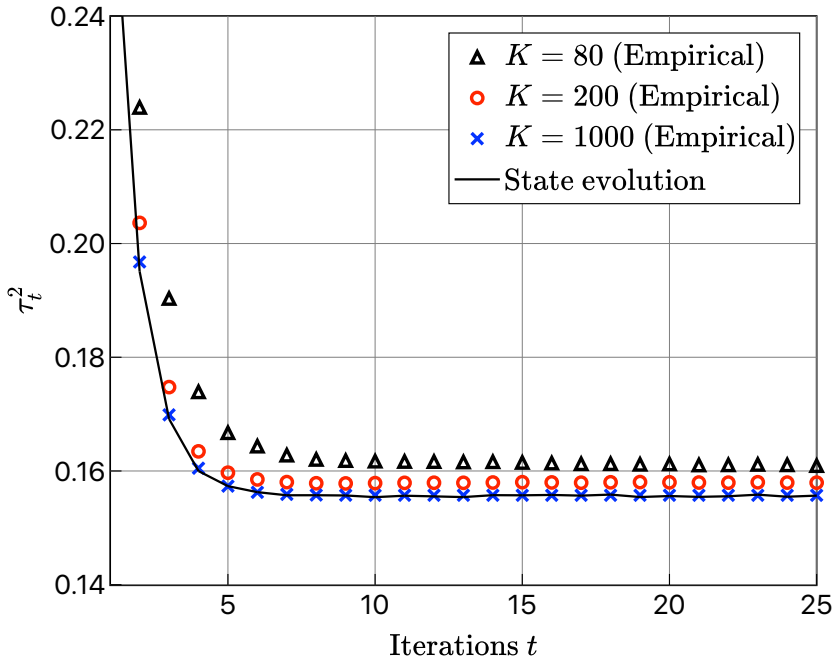

Fig. 2.3 The prediction of τ_t^2 via state evolution where $\omega = K/L = 4$, $\epsilon = K_{\rm a}/K = 0.1$, $M = 1$, $J = 6$, and SNR is 10 dB.

In this chapter, we hereafter focus on the case that $\beta_k = 1$, $\forall k$ to simplify the performance analysis, enabling to omit the expectation operations over β. This scenario can be realized by the compensation owing to the perfect power control and was considered in the conventional studies, *e.g.*, [67, 65, 69].

Fig. 2.3 shows the prediction of τ_t^2 via state evolution and the empirical one with MMV-AMP obtained by simulations, where $\omega = K/L = 4$, $\epsilon = K_{\rm a}/K = 0.1$, $M = 1$, $J = 6$, and SNR is 10 dB. We evaluate the performance for the different problem sizes of $K = 80$, 200, and 1000. Each entry of the measurement matrix $\mathbf{A}$ obeys i.i.d. complex Gaussian distribution with zero mean and variance $1/L$. As seen from the figure, the empirical performance is close to the prediction with state evolution when $200 \leq K$, especially $K = 1000$.

The statistical property of $\|(\mathbf{R}^{(t)})^{\rm T}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}\|_2^2$ is significant for the proposed threshold design. Notice that the entries of $(\mathbf{R}^{(t)})^{\rm H}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}$ can be modeled as Gaussian random variables owing to the equivalent signal model $\bar{\boldsymbol{X}}_{t,k} = \boldsymbol{X}_k + \tau_t \boldsymbol{V}_k$ in the state evolution. Thus, we here derive the PDF of $\|(\mathbf{R}^{(t)})^{\rm T}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}\|_2^2$.

Let us consider the case of non-active users. For $k \notin \mathcal{A}$, all elements of $\mathbf{x}_k^{(t)}$ become zero, and the term $\mathbf{x}_k^{(t)}$ is then negligible. On the other hand, the term of $(\mathbf{R}^{(t)})^{\rm T}\mathbf{a}_k^*$ corresponds to the random vector $\tau_t \mathbf{V}_k$, implying that $(\mathbf{R}^{(t)})^{\rm T}\mathbf{a}_k^*$ follows

$\mathcal{CN}(\mathbf{0}_{M(J+1)}, \tau_t^2 \mathbf{I}_{M(J+1)})$. Let ι_k be the user activity indicator for user k:

$$\iota_k = \begin{cases} 1 & (k \in \mathcal{A}) \\ 0 & (k \notin \mathcal{A}) \end{cases}. \tag{2.28}$$

Then, the PDF of $\|\bar{\mathbf{X}}_{t,k}\|_2^2$ for $n \notin \mathcal{A}$ is given by

$$p(\bar{X}|\iota_k = 0) = \frac{\tau_t^{-2M(J+1)} \bar{X}^{M(J+1)-1}}{\Gamma(M(J+1))} e^{-\frac{\bar{X}}{\tau_t^2}}, \tag{2.29}$$

where $\Gamma(\cdot)$ denotes the Gamma function.

Next, we consider the case of active users. In order to further simplify the analysis, we approximate $\|(\mathbf{R}^{(t)})^{\mathrm{T}} \mathbf{a}_k^* + \mathbf{x}_k^{(t)}\|_2^2$, as follows:

$$\|(\mathbf{R}^{(t)})^{\mathrm{T}} \mathbf{a}_k^* + \mathbf{x}_k^{(t)}\|_2^2 \approx \|(\mathbf{R}^{(t)})^{\mathrm{T}} \mathbf{a}_k^*\|_2^2 + \|\mathbf{x}_k^{(t)}\|_2^2. \tag{2.30}$$

It is obvious from the discussion for the case of non-active users that the PDF of the first term in (2.30) is equivalent to (2.29). In contrast, the PDF of the second term in (2.30) can be given by

$$p(S) = \frac{(J+1)^{-M} S^{M-1}}{\Gamma(M)} e^{-\frac{S}{J+1}}, \tag{2.31}$$

since $\|\mathbf{x}_k^{(t)}\|_2^2$ for $k \in \mathcal{A}$ can be regarded as the scaled squared-norm of the channels, i.e., $\|\mathbf{x}_k^{(t)}\|_2^2 = (J+1)\|\mathbf{h}_k\|_2^2$, because of the structure of the transmitted frame. Therefore, the distribution of (2.30) is approximately given by the distribution of the sum of two independent and non-identical Gamma random variables, *i.e.,* $X_1 \sim \Gamma(M(J+1), \tau_t^{-2})$ and $X_2 \sim \Gamma(M, (J+1)^{-1})$, where $\Gamma(\alpha, \beta)$ denotes the Gamma distribution. The PDF can thus be given by [81]

$$\begin{aligned} p(\bar{X}|\iota_k = 1) = {} & \frac{(J+1)^{-M} \tau_t^{-2M(J+1)}}{\Gamma(M(J+2))} \bar{X}^{M(J+2)-1} e^{-\frac{\bar{X}}{\tau_t^2}} \\ & \times {}_1\mathrm{F}_1\left(M; M(J+2); \left(\frac{1}{\tau_t^2} - \frac{1}{J+1}\right)\bar{X}\right), \end{aligned} \tag{2.32}$$

where ${}_1\mathrm{F}_1(\cdot;\cdot;\cdot)$ is a *confluent hypergeometric function* [81, 82]. Hence, the probability of MD can be approximately obtained by

$$P_{\mathrm{a}}^{\mathrm{MD}}(M) = \int_0^{\theta_{t,k}} p(\bar{X}|\iota_k = 1)d\bar{X}. \tag{2.33}$$

Finally, the threshold in (2.15) is designed based on the PDFs of (2.29) and (2.32). This threshold can be derived from the following LLR of the two situations:

$$\ln\left[\frac{p(\bar{X}|\iota_k = 1)}{p(\bar{X}|\iota_k = 0)}\right] \underset{\iota_k=0}{\overset{\iota_k=1}{\gtrless}} 0. \tag{2.34}$$

Hence, the proposed threshold is equivalent to the value of $\bar{X}$ where the above LLR becomes zero. Although this threshold cannot be expressed as the closed form, it can be obtained numerically.

Here, we would like to remark the threshold design in the conventional scheme [54]. This design is based on the fact that the PDFs of $(\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}$ given that the user is inactive or active are given by

$$p((\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}|\iota_k = 0) = \frac{\exp(-\|(\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}\|_2^2\tau_t^{-2})}{\pi^{M(J+1)}\tau_t^{2M(J+1)}}, \tag{2.35}$$

$$p((\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}|\iota_k = 1) = \frac{\exp(-\|(\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}\|_2^2(1+\tau_t^2)^{-1})}{\pi^{M(J+1)}(1+\tau_t^2)^{M(J+1)}}, \tag{2.36}$$

respectively. Consequently, the value of threshold $\theta_{t,k}$ in (2.15) is derived from the following criterion:

$$\ln\left[\frac{p((\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}|\iota_k = 1)}{p((\mathbf{R}^{(t)})^{\mathrm{T}}\mathbf{a}_k^* + \mathbf{x}_k^{(t)}|\iota_k = 0)}\right] \underset{\iota_k=0}{\overset{\iota_k=1}{\gtrless}} 0. \tag{2.37}$$

Figure 2.4 shows the relation between values of $\theta_{t,k}$ given in (2.37) and (2.34). In this figure, we set $J = 6$, $\omega = K/L = 4$, and $\epsilon = K_{\mathrm{a}}/K = 0.1$. As seen from the figure, the value of the proposed threshold is smaller than that of the conventional one, and the gap between them grows gradually with the increase in SNR. These facts imply that the AUD exploiting the proposed threshold tends to detect more users as active, lowering the probability of MD.

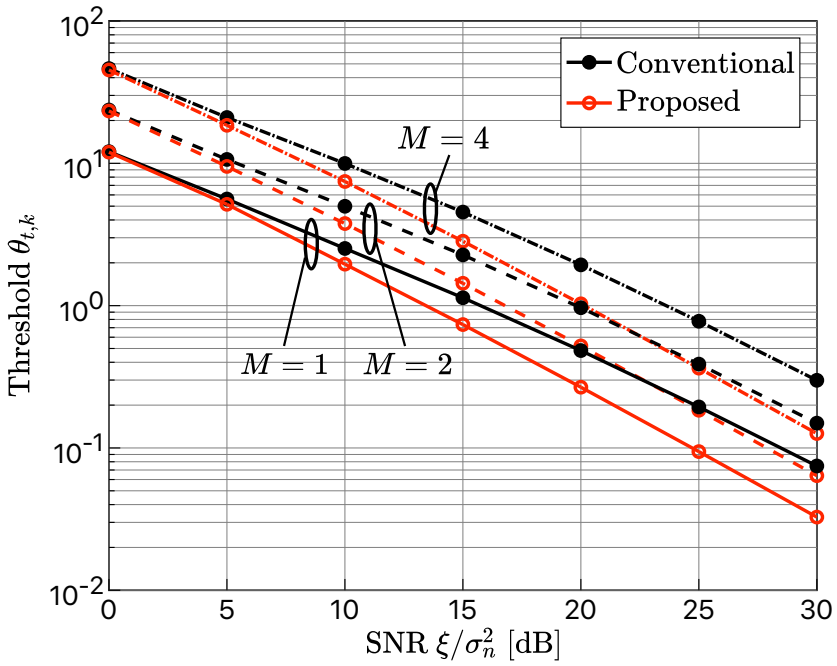

Fig. 2.4 Threshold based on [54] and the proposed threshold where $J = 6$, $\omega = K/L = 4$, and $\epsilon = K_{\mathrm{a}}/K = 0.1$.

2.3.2 Boosted AMP

In this subsection, we discuss the performance of the receiver using Boosted AMP under the situation that $M_{\mathrm{B}} < M(J+1)$ and $M_{\mathrm{B}}T_{\mathrm{rem}} \leq M(J+1)$, lowering the complexity than the one based on MMV-AMP.

According to the theoretical analyses in [54], the estimation error of MMV-AMP decreases with an increase in the number of measurements $M(J+1)$, resulting in the accurate AUD. This simultaneously implies that the performance of MMV-AMP in step 2 of Boosted AMP is inferior to the one solving (2.1) when $M_{\mathrm{B}} < M(J+1)$. Moreover, even if the receiver based on Boosted AMP repeats its process T_{rem} times, this cannot outperform the one with MMV-AMP that solves the original MMVR problem in (2.1) when $M_{\mathrm{B}}T_{\mathrm{rem}} \leq M(J+1)$.

In the receiver based on Boosted AMP, the performance of CE and MUD depends on the result of AUD because it exploits the MMSE detector in (2.21), which needs the matrix $\tilde{\mathbf{A}}$. If $|\hat{\mathcal{A}}| > L$, the detector has to solve an underdetermined problem and suffers from significant performance degradation. Meanwhile, the receiver based on MMV-AMP does not face such a problem because it simultaneously recovers data symbols for the whole user in the system. Hence, it is crucial to limit the candidates of the estimated active users correctly by the AUD using Boosted AMP. Further, it may

Table 2.2 Simulation parameters.

Number of users K	200
Number of active users K_{a}	20
Length of spreading sequences L	50
Number of data symbols J	6
Number of maximum iterations of MMV-AMP	100

be difficult for the receiver based on Boosted AMP to obtain the performance gain owing to the redesign of the criterion for AUD since the redesign might increases the probability of false alarm (FA). In light of the above, we conclude that the criterion in (2.20) is a judicious choice to perform CE and MUD.

2.4 Numerical Results

In this section, we evaluate the normalized mean-squared error (NMSE), SER, and the probability of the MD of our proposed receivers via computer simulations. We focus on the case that $\beta_k = 1,\ \forall k$, and randomly set K_{a} users as active users. For all simulations, the matrix $\mathbf{A}$ is obtained by uniformly randomly selecting the rows of a discrete Fourier transform (DFT) matrix, and each element of the random matrix $\mathbf{V}_i$ for Boosted AMP is assumed to take either $-1/\sqrt{M(J+1)}$ or $1/\sqrt{M(J+1)}$. Also, quadrature phase-shift keying (QPSK) is exployed. Moreover, the simulation parameters are set based on [65] and listed in TABLE 2.2. For simplicity, we use the notation "Boosted AMP $(T_{\mathrm{rem}}, M_{\mathrm{B}})$" to denote Boosted AMP where the number of iterations of ReMBo and the dimensions of the converted MMVR are T_{rem} and M_{B}, respectively.

2.4.1 NMSE Performance

First, we investigate the NMSE performance to confirm the accuracy of CE. The NMSE is defined as

$$\mathrm{NMSE} \triangleq \mathbb{E}\left[\frac{\|\mathbf{X}_{\mathrm{p}} - \hat{\mathbf{X}}_{\mathrm{p}}\|_{\mathrm{F}}^2}{\|\mathbf{X}_{\mathrm{p}}\|_{\mathrm{F}}^2}\right], \tag{2.38}$$

where $\mathbf{X}_{\mathrm{p}}$ denotes the period corresponding to the pilot period of $\mathbf{X}$ in (2.1) and $\hat{\mathbf{X}}_{\mathrm{p}}$ is the estimate of $\mathbf{X}_{\mathrm{p}}$. Furthermore, to provide a benchmark for performance comparison, we consider the oracle MMSE by assuming that the BS knows the active users perfectly.

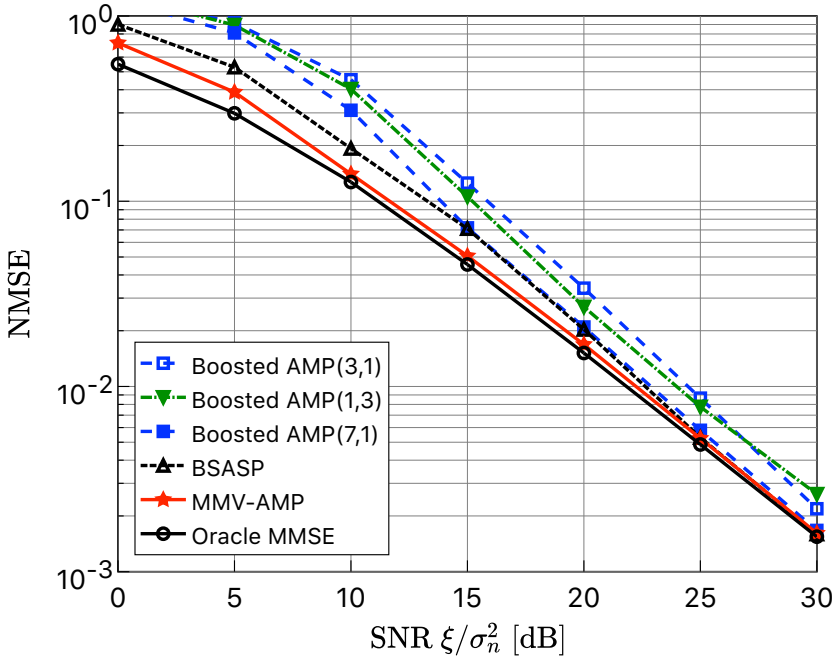

Fig. 2.5 The NMSE performances of the MD of BSASP, MMV-AMP, and Boosted AMP where $J = 6$, $K = 200$, $K_{\rm a} = 20$, $L = 50$, and $M = 1$.

This is because oracle MMSE becomes the lower bound of the performances of MMV-AMP and Boosted AMP with the MMSE denoiser.

Fig. 2.5 shows the NMSE performances of MMV-AMP, Boosted AMP, BSASP, and oracle MMSE where $M = 1$. Note that the performance of BSASP is obtained using scaled thresholds of [65] at each SNR, *i.e.*, $P_{\rm th} := L \cdot P_{\rm th}$, because of the difference in the definition of SNR. These results show that the receiver based on MMV-AMP can nearly achieve the accuracy of CE using oracle MMSE and outperform the conventional scheme of [65]. Besides, the performance of the receiver based on Boosted AMP solving $T_{\rm rem} = M(J+1)$ SMVR problems is inferior to that of the conventional one when SNR is less than 15 dB, while its performance is comparable when the SNR is more 15 dB. Although it is a natural result, the performances of the receiver with $(T_{\rm rem}, M_{\rm B}) = (3, 1)$ and $(T_{\rm rem}, M_{\rm B}) = (1, 3)$ are worse compared to that of other receivers, and they are inferior to the oracle MMSE by about 2 dB at NMSE $= 10^{-2}$.

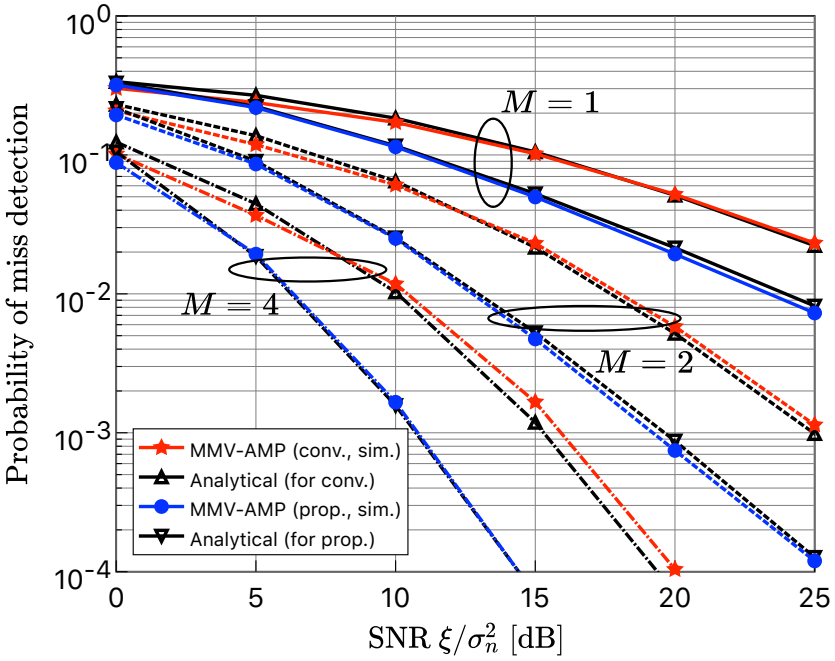

Fig. 2.6 The probabilities of MD of MMV-AMP with the conventional and proposed threshold where $J = 6$, $K = 200$, $K_{\rm a} = 20$, and $L = 50$. The abbreviations "conv.," "prop.," and "sim." denote "conventional," "proposed," and "simulation," respectively.

2.4.2 AUD Performance

Second, we investigate the accuracy of AUD through comparisons of the probabilities of MD as defined in (2.24) and those of FA given by

$$P^{\rm FA}(M) \triangleq \mathbb{E}\left[\frac{1}{K - K_{\rm a}}\left|\hat{\mathcal{A}} \setminus \mathcal{A}\right|\right]. \tag{2.39}$$

To investigate the effect of the threshold design, Fig. 2.6 depicts the probabilities of the MDs of MMV-AMP with the conventional and proposed thresholds. As the associated values, the analytical probabilities of MD $P_{\rm a}^{\rm MD}(M)$ in (2.33) are also depicted. In this figure, the abbreviations "conv.," "prop.," and "sim." denote "conventional," "proposed," and "simulation," respectively. As shown in Fig. 2.6, the MMV-AMP using the threshold based on (2.34) remarkably outperforms the conventional one, *i.e.*, using the threshold based on (2.37). The gain yields an SER performance that is close to that of the MMSE detector under the ideal assumption about AUD. Furthermore, Fig. 2.6 indicates that $P_{\rm a}^{\rm MD}(M)$ in (2.33) can predict the actual probability of MD approximately. In contrast, the probabilities of the FAs of MMV-AMP with the conventional and proposed thresholds are shown in Fig. 2.7. We can see that the

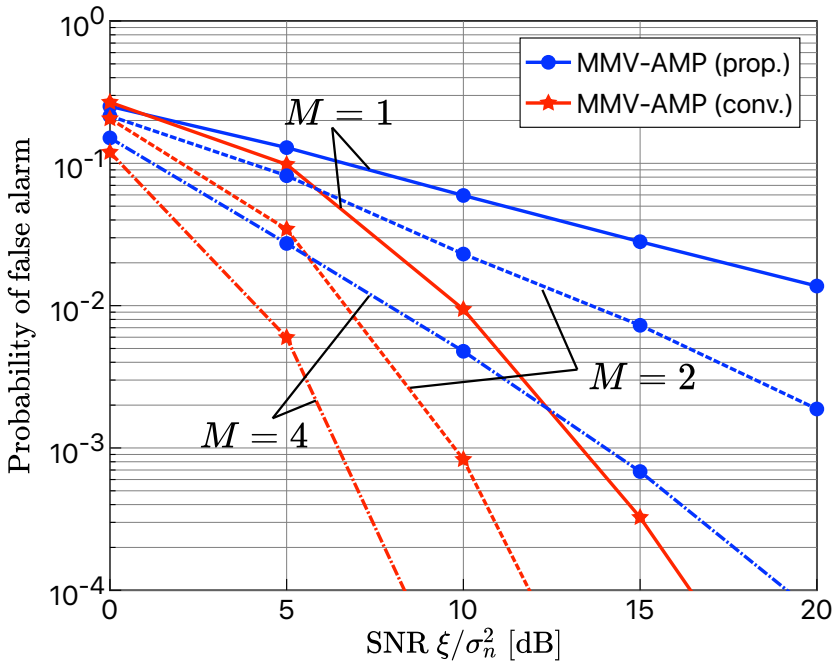

Fig. 2.7 The probabilities of FA of MMV-AMP with the conventional and proposed threshold where $J = 6$, $K = 200$, $K_{\mathrm{a}} = 20$, and $L = 50$.

proposed threshold is inferior to the conventional one in terms of the probability of FA, implying the tradeoff between MD and FA.

We also investigate the impact of sequence length on the AUD performance. Fig. 2.8 shows the probabilities of MD and FA for MMV-AMP with the conventional threshold, where $J = 6$, $K = 200$, $K_{\mathrm{a}} = 20$, and SNR $= 10$ dB. As seen from the figure, the performance gain obtained from the use of multiple antennas at the BS increases with the sequence length L, whereas it is small when L is close to K_{a}.

The probabilities of MD of the conventional and proposed receivers are shown in Fig. 2.9, where $M = 1$, 2, and 4. The probability of MD with the proposed threshold is shown as the result of MMV-AMP. For the evaluation of the receiver based on Boosted AMP, we set T_{rem} and M_{B} to be equal or less than $M(J+1)/2$. MMV-AMP can be observed to significantly outperform BSASP and to enhance the performance with an increasing M. In addition, the performance of Boosted AMP is comparable to that of BSASP, unlike the NMSE performance. The performance gap between Boosted AMP and MMV-AMP becomes smaller with an increase in M, while a large gap remains even when $M = 4$, *e.g.*, about 7 dB at $P^{\mathrm{MD}}(M) = 10^{-4}$. This is a reparation to significantly reduce the computational complexity. In contrast, the performance gap between the Boosted AMP using AMP and the one using MMV-AMP enlarges with

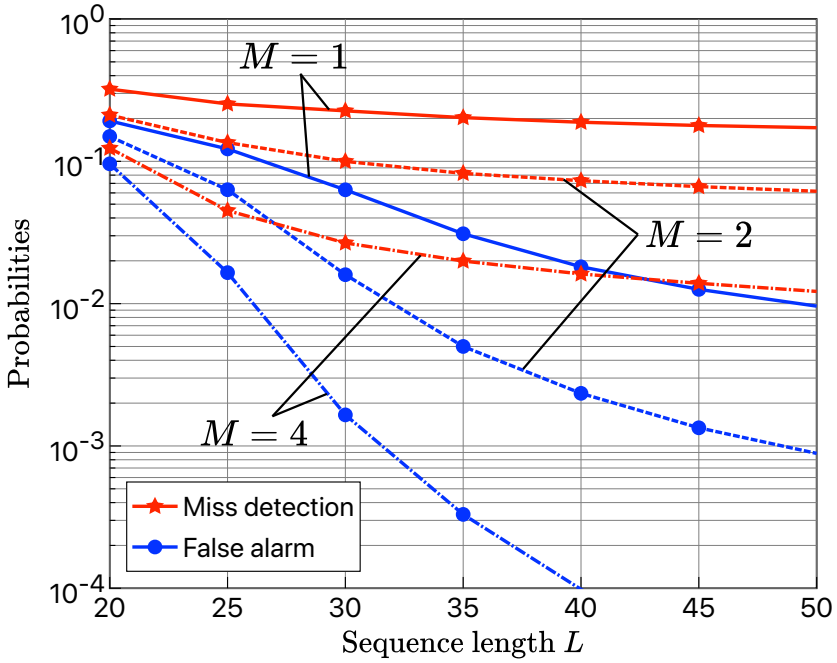

Fig. 2.8 Impact of sequence length L on the AUD performance of MMV-AMP with the conventional threshold where $J = 6$, $K = 200$, $K_{\mathrm{a}} = 20$, and SNR $= 10$ dB.

an increasing M. This implies that the tradeoff between performance and complexity is affected by the increase in the number of measurements at the BS.

2.4.3 SER Performance

Finally, we evaluate the SER performance of our proposed receivers. The symbol error is defined as the event that the AUD for the actual active user fails or the data symbol is not detected even if the AUD succeeds.

To begin with, Fig. 2.10 depicts the SER performances of oracle MMSE and MMV-AMP with the conventional and proposed thresholds for $M = 1$, 2, and 4. The performance of BSASP is also shown as the benchmark for $M = 1$. It is observed that MMV-AMP can nearly achieve the SER performance of oracle MMSE by using the threshold based on (2.34), unlike conventional studies, *i.e.*, [65, 67]. Moreover, the growth of the performance gap between MMV-AMP and oracle MMSE in the high SNR regime decreases owing to the improvement of the probability of MD. Although it is obvious from the result shown in Fig. 2.6, the gap gradually degrades with the growth of M, as observed in Fig. 2.10. As a result, by utilizing the properly designed threshold, the performance of MMV-AMP can be improved as contrasted with a straightforward implementation based on [54]. In addition to the evaluation of the probability of MD,

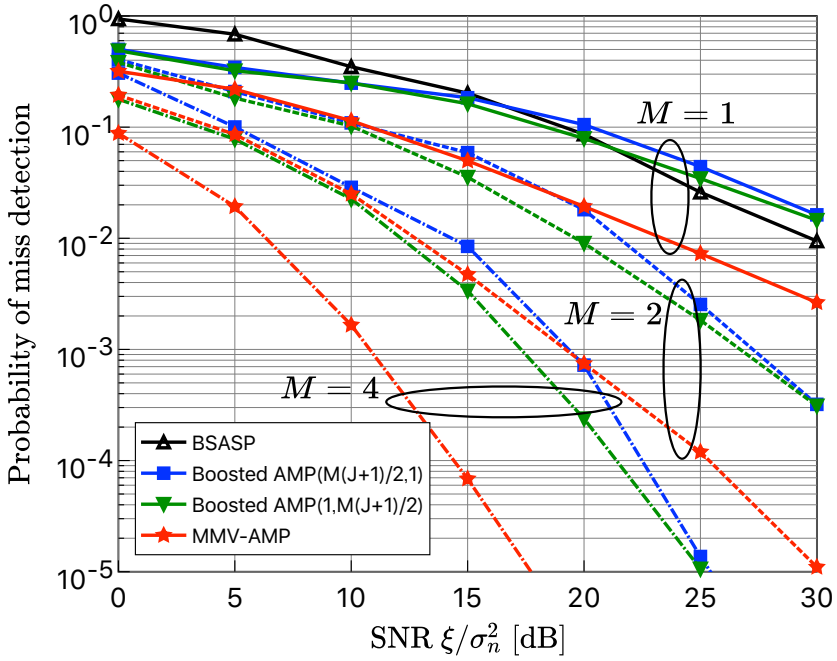

Fig. 2.9 The probabilities of MD of BSASP, MMV-AMP, and Boosted AMP where $J = 6$, $K = 200$, $K_{\rm a} = 20$, and $L = 50$.

the SER performance with the proposed threshold is hereafter shown as a result of MMV-AMP.

Next, we discuss the effect of the threshold for AUD on the performance of the receiver based on Boosted AMP. Fig. 2.11 shows the numerical example of the SER performance of the receiver using Boosted AMP versus the value of threshold where $M = 2$, $M_{\rm B} = 1$, $T_{\rm rem} = 7$, and SNR=20 dB. The figure also depicts the associated probabilities of MD and FA. The main aim of redesigning the threshold is to minimize the SER performance, while it is observed that the performance based on the criterion in (2.20) is almost the optimal one. As mentioned in Section 2.3.2, this is because the receiver based on Boosted AMP utilizes the MMSE detector in (2.21) to estimate the channels and data symbols, and its performance depends on both MD and FA, *i.e.*, the size of $\tilde{\mathbf{A}}$ in (2.21). Therefore, redesigning the criterion of AUD like that for the MMV-AMP does not work in Boosted AMP.

Third, the performances of our proposals and oracle MMSE with $M = 1$, 2, and 4 are shown in Fig. 2.12, which includes the SER performance of BSASP as the benchmark for $M = 1$. As shown in the figure, the performance gap between Boosted AMP and MMV-AMP becomes smaller, *i.e.*, about 6 dB at SER $= 10^{-4}$ for $M = 4$, as compared with the result shown in Fig. 2.9. On the other hand, as the value of

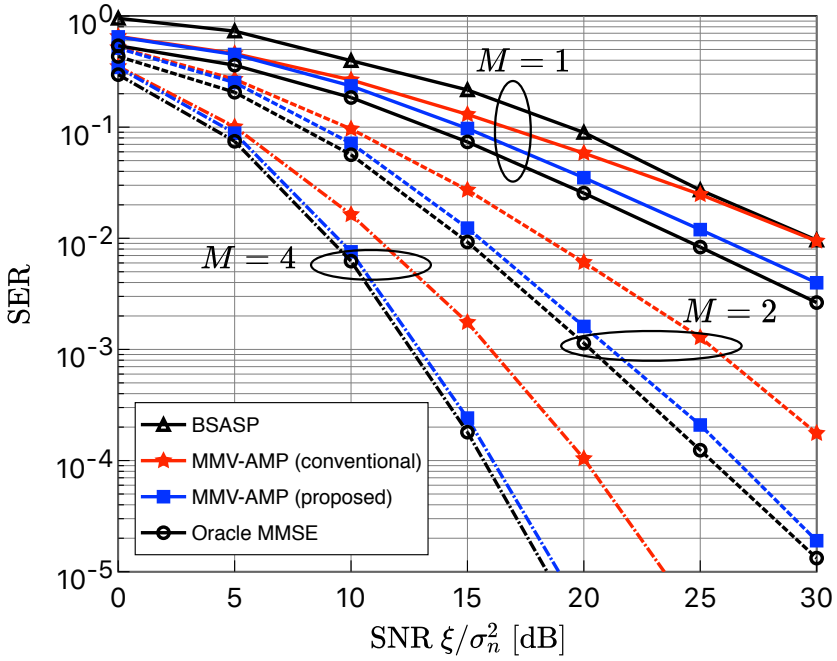

Fig. 2.10 The SER performances of MMV-AMP with the conventional and proposed threshold where $J = 6$, $K = 200$, $K_{\rm a} = 20$, and $L = 50$. As the benchmark for $M = 1$, the SER performance of BSASP is also shown.

M increases, the performance gap between the receiver based on Boosted AMP for solving multiple SMVR problems and the one solving the MMVR problem increases.

Finally, we discuss the effect of how to reduce the complexity in the receiver based on Boosted AMP, and we focus on the case of $M = 4$. Fig. 2.13 shows the SER performance of MMV-AMP, Boosted AMP, and oracle MMSE for $M = 4$. Note that the receiver based on MMV-AMP with the conventional threshold is considered. Here, we consider the pairs of $M_{\rm B}$ and $T_{\rm rem}$, which satisfy $M_{\rm B}T_{\rm rem} = M(J+1)/4$ or $M_{\rm B}T_{\rm rem} = M(J+1)/2$. The receiver based on Boosted AMP(14,1) solves 14 SMVR problems, in other words, it is equivalent that Boosted AMP(14,1) utilizes 14 measurement vectors. As shown in Fig. 2.13, however, the performance of Boosted AMP(14,1) is comparable with that of Boosted AMP(1,7), despite the fact that the latter solves the MMVR problem composed of 7 measurement vectors. This implies that the degradation of SER performance due to the reduction of complexity escalates when the original MMVR problem is converted into the SMVR problem. Moreover, Boosted AMP(1,14) can achieve the performance gap within 2 dB from MMV-AMP at SER=10^{-4} while the former uses only half of the dimensions compared with the latter. Hence, the receiver based on Boosted AMP using MMV-AMP, *i.e.*, $M_{\rm B} > 1$, would

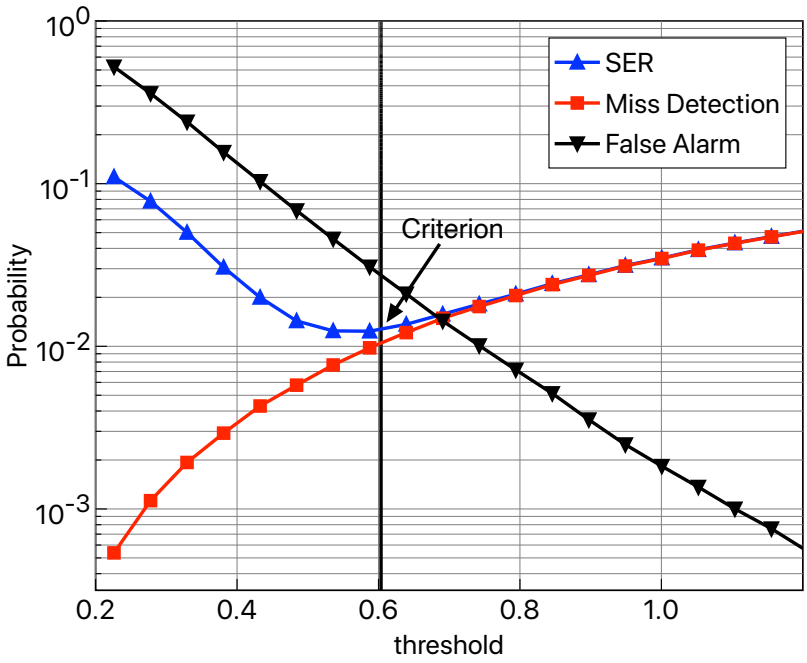

Fig. 2.11 Numerical example of the performances of Boosted AMP versus the value of the threshold.

have the potential to satisfy the tradeoff between the performance and complexity for a large M.

2.5 Chapter Summary

In this chapter, we investigated a narrowband GF-NOMA system with a multiple-antenna BS and proposed two low-complexity receivers. One of our proposals exploits MMV-AMP to perform AUD, CE, and MUD jointly, and its analytical performance was derived. Besides, we provided the threshold for AUD based on the theoretical analysis. The other proposal is based on Boosted AMP, which is the combination of MMV-AMP and ReMBo, to further reduce the computational complexity.

This chapter revealed that our proposed receivers can attain performance superior or comparable to that of the state-of-the-art BSASP while lowering the computational complexity. It is worth noting that the SER performance of the receiver, which exploits MMV-AMP and the theoretically-designed threshold, can approach that of the grant-based code-domain NOMA with linear MMSE detector. Hence, we conclude that this chapter largely contributes to the design of receivers for GF-NOMA systems to perform AUD, CE, and MUD efficiently.

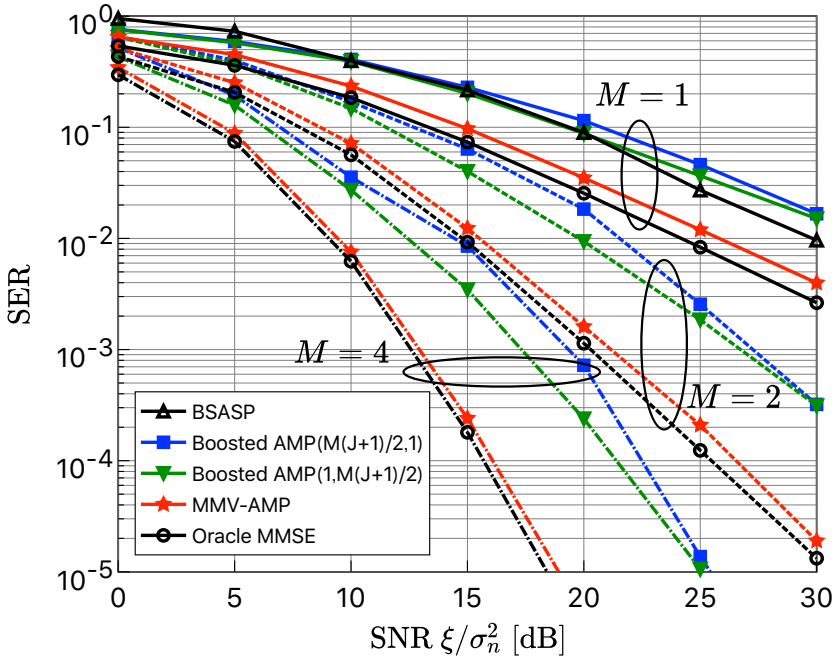

Fig. 2.12 The SER performance of our proposals and oracle MMSE where $J = 6$, $K = 200$, $K_{\mathrm{a}} = 20$, and $L = 50$. As the benchmark for $M = 1$, the SER performance of BSASP is also shown.

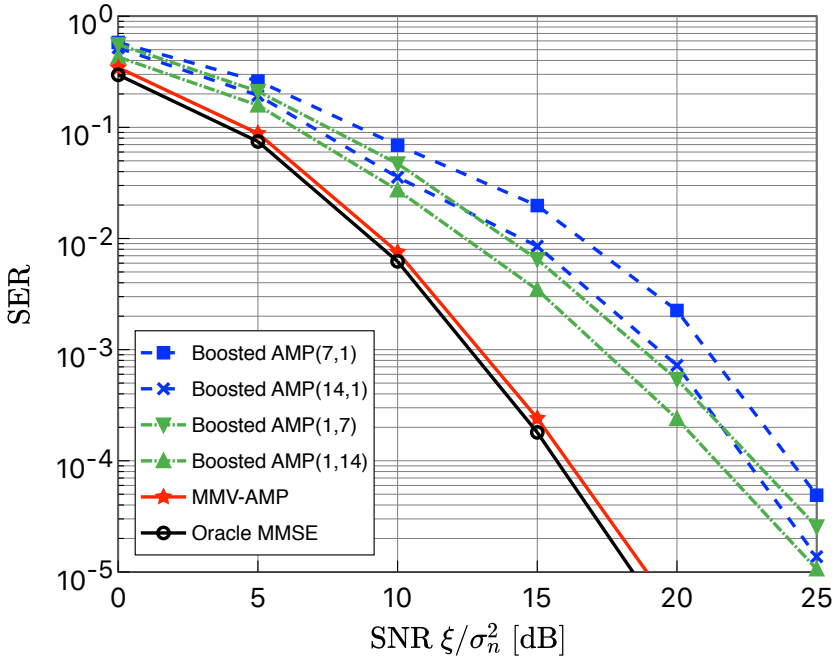

Fig. 2.13 The SER performance where $J = 6$, $K = 200$, $K_{\mathrm{a}} = 20$, $L = 50$, and $M = 4$.

Chapter 3

Receivers for GF-NOMA Against the Effects of Large-Scale Fading or CFOs

The previous chapter has shown that two proposed receivers for narrowband GF-NOMA systems with multiple-antenna BS can improve the SER performance while lowering the computational complexity. On the other hand, they can function only when the system can perform the perfect time and frequency synchronization and enable the BS to know the LSF coefficients of all users ideally. As mentioned in Section 1.6, although some approaches have been proposed under an imperfect frequency synchronization or ignorance of the LSF coefficients, they still have weak points in terms of practicality.

This chapter presents the solutions to the second and third issues mentioned in Section 1.6. We first consider an alternate scheme to perform AUD and CE when all users are time and frequency synchronized while the BS does not know the LSF coefficients. Then, we propose a promising scheme, which integrates the EM algorithm with the MMV-AMP algorithm, and an appropriate decision rule for AUD. Through computer simulations, it is demonstrated that the performance of the proposed scheme is comparable to that of MMV-AMP with the prior knowledge of LSF coefficients. After that, we investigate the AUD in the scenario that CFOs exist while perfectly compensating the LSF coefficients by power control. We obtain a tailored formulation inspired by an idea of array-signal processing and propose a covariance-based AUD using the CD method. Moreover, we consider the box-constraint obtained from the formulation. Numerical results demonstrate the proposed scheme can estimate active users accurately.

3.1 Overview of Related Works

As shown in Section 1.5, many approaches to cope with the fundamental challenges in GF-NOMA have been actively investigated. However, many of them consider the ideal scenarios, where all users are perfectly synchronized with a common BS or the BS can utilize the LSF coefficients of all users. In fact, prior knowledge of the wireless channels such as LSF coefficients would be unavailable at the BS ahead of the uplink grant-free transmission. Moreover, if the users utilize cheap crystal oscillators for reducing cost, like narrowband IoT (NB-IoT) devices, the CFOs between users and the BS would be inevitable, degrading the performance of GF-NOMA systems [83]. From a practical point of view, the above difficulties should hence be addressed to realize GF-NOMA systems, leading to the motivation of this chapter.

For the case that the BS does not know the knowledge of the channels, some promising approaches have been proposed [55, 56, 63]. In [63], the JACE based on the GMMV-AMP algorithm has been proposed. This method decouples the matrix estimation problem into scalar estimation problems and utilizes the element-wise message passing procedures, lowering the required computational complexity. In contrast, it does not take full advantage of the row sparsity that yields performance gain. As an alternative method, the JACE based on a convex optimization has been proposed in [55], which solves the optimization problem that incorporates a regularization reflecting the row sparsity. However, this scheme requires an exhaustive search of regularization parameters for accurate estimation. Moreover, in [56], the problem of AUD and CE was formulated into a low-rank sparse matrix recovery method and solved using a Riemannian optimization technique. However, this approach only works under the assumption that the BS is equipped with antennas more than the number of active users. In other words, more than a hundred antennas are required when a hundred users are active at each coherence time, like the scenario in [15]. This constraint is undesirable in terms of hardware cost.

On the other hand, in the presence of CFOs, the systems suffer from significant performance degradation due to the phase rotation that comes from them. From a mathematical point of view, the CFOs result in uncertainty in a measurement matrix, which degrades the performance of CS-based approaches [84], or the nonlinear coupling between CFOs and channel coefficients. Furthermore, these difficulties cause the extra non-convexity and ill-conditioned measurement matrix, making convex optimization-based approaches and message-passing algorithms such as AMP not applicable. Toward this end, an AUD for GF-NOMA systems under the CFOs, which is a combination of the Taylor expansion and BCD method, has been proposed in [53]. However, this

approach needs to know the maximum number of active users in advance, whereas the BS is unable to obtain such knowledge in practical systems due to the nature of GF-NOMA.

In light of the above, the receiver design against the scenarios in which CFOs exist or the LSF coefficients are unknown at the BS, is still challenging. Thus, this chapter addresses such a receiver design, and its contributions are summarized as follows.

JACE based on the EM-MMV-AMP Algorithm

We propose a new algorithm called EM-MMV-AMP to jointly perform AUD and CE without prior knowledge of the wireless channels, namely LSF coefficients. This algorithm is based on the MMV-AMP algorithm [54] integrated with the EM algorithm in a manner similar to that in [85]. Furthermore, we revisit the decision rule for AUD and propose an appropriate rule for the EM-MMV-AMP algorithm. Finally, we demonstrate the superior performance of AUD and CE based on EM-MMV-AMP via computer simulations.

Covariance-based AUD in the Presence of CFOs

We propose a new approach to perform AUD where CFOs exist. The proposed approach utilizes the reformulation based on the *peak-detection* technique considered for direction-of-arrival estimation problems [86] so as to enhance the sparsity. Moreover, we employ the CD method such as ones proposed in [51], which can overcome the non-convexity due to the coupling between CFOs and channels, avoiding the impractical prior knowledge for AUD, *i.e.*, the maximum number of active users. Computer simulations show that our proposal can detect the activity pattern accurately.

3.2 System Model

This chapter considers the same uplink GF-NOMA system as Fig. 2.1, whereas the scenario differs from Chapter 2. Concretely, we focus on the pilot part of the uplink transmission, which corresponds to the case that $J = 0$ in Chapter 2, and reformulate the signal model including CFOs. Assuming that each single-antenna user is equipped with a low-cost crystal oscillator, the CFO between the user and the common BS arises. Then, the angular frequency caused by the CFO is modeled as $\varpi_k \triangleq 2\pi\Delta f_k T_\mathrm{s}$, where Δf_k and T_s are the frequency offset in Hz and the sampling period, respectively.

Furthermore, we assume that all users are time-synchronized and $K_{\mathrm{a}} \leq K$ users are active to transmit in coherence time consisting of L time slots.

In the same manner as Chapter 2, we define the total transmission power, the set of active users, the unique sequence, and the channel vector as ξ, $\mathcal{A} \subset \{1, \ldots, K\}$, $\mathbf{a}_k = [a_{k,1}, \ldots, a_{k,L}]^{\mathrm{T}} \in \mathbb{C}^{L \times 1}$, and $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}_M, \beta_k \mathbf{I}_M)$, respectively. Besides, the cardinality of $\mathcal{A}$ is K_{a}, and β_k is the LSF coefficient of user k. Then, the received signals at the BS in the ℓ-th time slot is given by

$$\mathbf{y}_\ell = \sqrt{\xi} \sum_{k \in \mathcal{A}} \mathrm{e}^{j(\ell-1)\varpi_k} a_{k,\ell} \mathbf{h}_k + \mathbf{z}_\ell \in \mathbb{C}^{M \times 1}, \quad \ell = 1, \ldots, L, \tag{3.1}$$

where $\mathrm{e}^{j(\ell-1)\varpi_k}$, $a_{k,\ell}$, and $\mathbf{z}_\ell \sim \mathcal{CN}(\mathbf{0}_M, \sigma_{\mathrm{n}}^2 \mathbf{I}_M)$ represent the phase rotation at the ℓ-th time slot due to the angular frequency ϖ_k, the ℓ-th element of $\mathbf{a}_k$, and the noise vector, respectively. Moreover, the received signals over L time slots, $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_L]^{\mathrm{T}} \in \mathbb{C}^{L \times M}$, can be expressed as

$$\begin{aligned}
\mathbf{Y} &= \sqrt{\xi} \sum_{k \in \mathcal{A}} \begin{bmatrix} (1 \cdot a_{k,1}) \mathbf{h}_k^{\mathrm{T}} \\ (\mathrm{e}^{j\varpi_k} \cdot a_{k,2}) \mathbf{h}_k^{\mathrm{T}} \\ \vdots \\ (\mathrm{e}^{j(L-1)\varpi_k} \cdot a_{k,L}) \mathbf{h}_k^{\mathrm{T}} \end{bmatrix} + [\mathbf{z}_1, \ldots, \mathbf{z}_L]^{\mathrm{T}} \\
&= \sqrt{\xi} \sum_{k \in \mathcal{A}} \begin{bmatrix} a_{k,1} & 0 & \cdots & 0 \\ 0 & a_{k,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{k,L} \end{bmatrix} \begin{bmatrix} 1 \\ \mathrm{e}^{j\varpi_k} \\ \vdots \\ \mathrm{e}^{j(L-1)\varpi_k} \end{bmatrix} \mathbf{h}_k^{\mathrm{T}} + \mathbf{Z} \\
&= \sqrt{\xi} \sum_{k \in \mathcal{A}} \mathrm{diag}(\mathbf{a}_k) \mathbf{b}_L(\varpi_k) \mathbf{h}_k^{\mathrm{T}} + \mathbf{Z},
\end{aligned} \tag{3.2}$$

where $\mathbf{b}_L(\varpi_k) \triangleq [1, \mathrm{e}^{j\varpi_k}, \ldots, \mathrm{e}^{j(L-1)\varpi_k}]^{\mathrm{T}} \in \mathbb{C}^{L \times 1}$. It is worth noting that if a frequency synchronization is perfectly performed, (3.2) is equivalent to (2.1) with $J = 0$. In such a situation, we can obtain the relation: $\mathrm{diag}(\mathbf{a}_k)\mathbf{b}_L(\varpi_k)|_{\varpi_k=0} = \mathrm{diag}(\mathbf{a}_k)\mathbf{1}_L = \mathbf{a}_k$.

3.3 EM-MMV-AMP Based Approach

In this section, we consider the scheme to performAUD and CE without the prior knowledge of LSF coefficients. Furthermore, we focus on the scenario in which all users are time and frequency synchronized ideally, that is, $\varpi_k = 0$ for $k = 1. \ldots, K$ in (3.2).

Hence, we consider the following signal model:

$$\mathbf{Y} = \sqrt{\xi}\mathbf{A}\mathbf{X} + \mathbf{Z}, \tag{3.3}$$

where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_k] \in \mathbb{C}^{L\times K}$ and $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_k]^{\mathrm{T}} \in \mathbb{C}^{K\times M}$ with

$$\mathbf{x}_k = \begin{cases} \mathbf{h}_k, & k \in \mathcal{A}, \\ \mathbf{0}_M, & \text{otherwise.} \end{cases} \tag{3.4}$$

In this section, β_k is modeled as $\beta_k = -128.1 - 37.6\log_{10}(d_k)[\mathrm{dB}]$, where d_k denotes the distance measured in km between BS and user k.

We first presents an overview of the related scheme, GMMV-AMP [64], to clarify the difference between the conventional and proposed approaches. We then describe the detail of the proposed algorithm, EM-MMV-AMP, and introduce the proper criterion for AUD. Finally, we demonstrate superior performance of the proposed approach via computer simulations.

3.3.1 Overview of the GMMV-AMP Algorithm

The problem of AUD and CE can be regarded as the matrix estimation problem in (3.3). GMMV-AMP decouples the estimation problem into KM scalar estimation problems under the assumption of a prior distribution of $\mathbf{X}$ in (3.3) [85], given by

$$p(\mathbf{X}) = \prod_{m=1}^{M}\prod_{k=1}^{K}[(1-\gamma_{k,m})\delta_0(x_{k,m}) + \gamma_{k,m}\mathcal{CN}(\mu,\tau)], \tag{3.5}$$

where $\gamma_{k,m} \in [0,1]$ presents the sparsity ratio. According to this assumption, this scheme hence executes the element-wise message passing procedure to obtain the MMSE estimate of $\mathbf{X}$, *i.e.*, the posterior mean.

Moreover, GMMV-AMP introduces the unknown hyperparameters, *i.e.*, $\mathbf{\Theta} = \{\mu, \tau, \sigma, \gamma_{k,m}, \forall k, m\}$, to be learned by the EM algorithm, avoiding the need for prior knowledge of the channels $\{\mu, \tau, \gamma_{k,m}, \forall n, m\}$ and the noise variance σ_{n}^2. As the proposed scheme in Chapter 5 exploits the GMMV-AMP algorithm, its detail will be described there.

Here, we mention two differences between GMMV-AMP and EM-MMV-AMP; 1) GMMV-AMP decouples the matrix estimation problem into scalar problems, and 2) GMMV-AMP requires updating the hyperparameter corresponding to the noise

variance additionally. Particularly, the first difference implies that GMMV-AMP does not take advantage of the row sparsity of $\mathbf{X}$, resulting in the performance degradation as compared to (EM-)MMV-AMP. This effect will be confirmed in Section 3.3.4 via computer simulations.

3.3.2 EM-MMV-AMP Algorithm

Unlike GMMV-AMP, EM-MMV-AMP executes the row-wise message passing procedure assuming the following prior distribution model:

$$p_X(\mathbf{x}_k;\epsilon,\beta_k) = (1-\epsilon)\delta_0(\mathbf{x}_k) + \epsilon\mathcal{CN}(\mathbf{0}_M,\beta_k\mathbf{I}_M), \quad \forall k, \tag{3.6}$$

where $\epsilon = K_{\mathrm{a}}/K \in [0,1]$ is the activity ratio, which is assumed to be available at the BS. This procedure is based on two following calculations:

$$\mathbf{x}_k^{(t+1)} = \eta_{t,k}\left((\mathbf{R}^{(t)})^{\mathrm{H}}\mathbf{a}_k + \mathbf{x}_k^{(t)}\right), \tag{3.7}$$

$$\mathbf{R}^{(t+1)} = \mathbf{Y} - \mathbf{A}\mathbf{X}^{(t+1)} + \frac{K}{L}\mathbf{R}^{(t)}\sum_{k=1}^{K}\frac{\eta'_{t,k}\left((\mathbf{R}^{(t)})^{\mathrm{H}}\mathbf{a}_k + \mathbf{x}_k^{(t)}\right)}{K}, \tag{3.8}$$

where we have

$$\tau_t^2 = \frac{\|\mathbf{R}^{(t)}\|_{\mathrm{F}}^2}{LM}. \tag{3.9}$$

Notice that although the above operations are the same as those of MMV-AMP described in Section 2.2.2, the dimension of the matrices are different, *i.e.*, $\mathbf{R}^{(t)} \in \mathbb{C}^{L\times M}$, $\mathbf{X}^{(t)} \in \mathbb{C}^{K\times M}$, because this chapter focuses on the pilot part, namely $J = 0$. Furthermore, we employ the MMSE denoiser[54] as the function $\eta_{t,k}(\cdot)$. Here, its input is denoted by $\hat{\mathbf{x}}_{t,k} \triangleq (\mathbf{R}^{(t)})^{\mathrm{H}}\mathbf{a}_k + \mathbf{x}_k^{(t)}$ and modeled as

$$\hat{\mathbf{x}}_{t,k} = \mathbf{x}_k + \tau_t\mathbf{v}_k, \ \forall k, \tag{3.10}$$

where $\mathbf{v}_k \sim \mathcal{CN}(\mathbf{0}_M,\mathbf{I}_M)$.

EM-MMV-AMP introduces the hyperparameters $\boldsymbol{\beta}_t = (\beta_{t,1},\ldots,\beta_{t,k})$ in order to avoid the need for prior information of LSF coefficients. By utilizing $\hat{\mathbf{x}}_{t,k}$ and τ_t^2 at the t-th iteration of MMV-AMP, we approximate the true marginal distribution, as follows

$$p_{X|Y}(\mathbf{x}_k|\mathbf{Y},\hat{\mathbf{x}}_{t,k},\tau_t^2,\boldsymbol{\beta}_t) \triangleq \frac{p_X(\mathbf{x}_k;\epsilon,\boldsymbol{\beta}_t)\mathcal{CN}(\hat{\mathbf{x}}_{t,k},\tau_t^2\mathbf{I}_M)}{\int_{\mathbf{x}} p_X(\mathbf{x};\epsilon,\boldsymbol{\beta}_t)\mathcal{CN}(\hat{\mathbf{x}}_{t,k},\tau_t^2\mathbf{I}_M)}, \tag{3.11}$$

where $p_X(\mathbf{x}_k;\epsilon,\boldsymbol{\beta}_t)$ is the prior distribution of $\mathbf{x}_k$ given by

$$p_X(\mathbf{x}_k;\epsilon,\boldsymbol{\beta}_t)=(1-\epsilon)\delta_0(\mathbf{x}_k)+\epsilon\mathcal{CN}(\mathbf{0}_M,\beta_{t,k}\mathbf{I}_M). \tag{3.12}$$

Plugging the prior model (3.12) into (3.11), the approximated posterior can be obtained by

$$p_{X|Y}(\mathbf{x}_k|\mathbf{Y},\hat{\mathbf{x}}_{t,k},\tau_t^2,\boldsymbol{\beta}_t)=(1-\phi_{t,k})\delta_0(\mathbf{x}_k)+\phi_{t,k}\mathcal{CN}(\boldsymbol{\mu}_{t,k},\nu_{t,k}\mathbf{I}_M), \tag{3.13}$$

where we have

$$\phi_{t,k}\triangleq\frac{1}{1+\frac{1-\epsilon}{\epsilon}\exp\left(-M(\pi_{t,k}-\psi_{t,k})\right)}, \tag{3.14}$$

$$\pi_{t,k}=\frac{1}{M}\left(\frac{1}{\tau_t^2}-\frac{1}{\beta_{t,k}+\tau_t^2}\right)\|\hat{\mathbf{x}}_{t,k}\|_2^2, \tag{3.15}$$

$$\psi_{t,k}=\ln\left(1+\frac{\beta_{t,k}}{\tau_t^2}\right), \tag{3.16}$$

$$\boldsymbol{\mu}_{t,k}\triangleq\frac{\beta_{t,k}}{\beta_{t,k}+\tau_t^2}\hat{\mathbf{x}}_{t,k}, \tag{3.17}$$

$$\nu_{t,k}\triangleq\frac{\beta_{t,k}\tau_t^2}{\beta_{t,k}+\tau_t^2}. \tag{3.18}$$

Note that (3.13) can be derived via the multiplication rule:

$$\begin{aligned}
&\mathcal{CN}(\mathbf{a},\alpha\mathbf{I}_M)\mathcal{CN}(\mathbf{b},\beta\mathbf{I}_M)\\
=&\frac{1}{(\pi\alpha)^M}\exp\left(-\frac{\|\mathbf{x}-\mathbf{a}\|_2^2}{\alpha}\right)\frac{1}{(\pi\beta)^M}\exp\left(-\frac{\|\mathbf{x}-\mathbf{b}\|_2^2}{\beta}\right)\\
=&\frac{1}{(\pi(\alpha+\beta))^M}\exp\left(-\frac{\|\mathbf{a}-\mathbf{b}\|_2^2}{\alpha+\beta}\right)\\
\times&\underbrace{\frac{1}{\pi^M\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)^{-M}}\exp\left(-\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)\left\|\mathbf{x}-\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)^{-1}\left(\frac{\mathbf{a}}{\alpha}+\frac{\mathbf{b}}{\beta}\right)\right\|_2^2\right)}_{\mathcal{CN}\left(\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)^{-1}\left(\frac{\mathbf{a}}{\alpha}+\frac{\mathbf{b}}{\beta}\right),\left(\frac{1}{\alpha}+\frac{1}{\beta}\right)^{-1}\mathbf{I}_M\right)}.
\end{aligned} \tag{3.19}$$

According to [85], the update rule of hyperparameters is given by

$$\boldsymbol{\beta}_{t+1}=\arg\max_{\boldsymbol{\beta}}\mathbb{E}\left[\ln p(\mathbf{X},\mathbf{Y};\boldsymbol{\beta})|\mathbf{Y};\boldsymbol{\beta}_t\right], \tag{3.20}$$

where $\mathbb{E}[\cdot|\mathbf{Y};\boldsymbol{\beta}_t]$ is the expectation conditioned on $\mathbf{Y}$ with the parameters $\boldsymbol{\beta}_t$.

We first focus on the update of β_k, given the previous parameters $\boldsymbol{\beta}_t$. The term $p(\mathbf{X}, \mathbf{Y}; \boldsymbol{\beta})$ in (3.20) can be expressed by $\tilde{\mathrm{C}} \prod_{k=1}^{K} p_X(\mathbf{x}_k, \mathbf{Y}; \epsilon, \beta_k)$, where a constant $\tilde{\mathrm{C}}$ is independent of $\beta_k, \forall k$. The update rule for β_k is hence given by

$$\beta_{t+1,k} = \underset{\beta_k > 0}{\arg\max}\ \mathbb{E}[\ln p_X(\mathbf{x}_k; \epsilon, \beta_k) | \mathbf{Y}; \boldsymbol{\beta}_t]. \tag{3.21}$$

The solution of (3.21) necessarily satisfies

$$\int_{\mathbf{x}_k} p_{X|Y}(\mathbf{x}_k | \mathbf{Y}; \boldsymbol{\beta}_t) \frac{d}{d\beta_k} \ln p_X(\mathbf{x}_k; \epsilon, \beta_k) = 0. \tag{3.22}$$

The derivative of the prior distribution given in (3.6) is

$$\begin{aligned} \frac{d}{d\beta_k} \ln p_X(\mathbf{x}_k; \epsilon, \beta_k) &= \left(\frac{\|\mathbf{x}_k\|_2^2}{\beta_k^2} - \frac{M}{\beta_k} \right) \frac{\epsilon \mathcal{CN}(\mathbf{0}_M, \beta_k \mathbf{I}_M)}{p_X(\mathbf{x}_k; \epsilon, \beta_k)} \\ &= \begin{cases} \frac{\|\mathbf{x}_k\|_2^2}{\beta_k^2} - \frac{M}{\beta_k}, & \mathbf{x}_k \neq \mathbf{0}_M, \\ 0, & \mathbf{x}_k = \mathbf{0}_M. \end{cases} \end{aligned} \tag{3.23}$$

By substituting (3.23) and (3.13), we can obtain the following update rule for β_k:

$$\beta_{t+1,k} = \frac{1}{M} \|\boldsymbol{\mu}_{t,k}\|_2^2 + \nu_{t,k}, \ \forall k. \tag{3.24}$$

The derivation of (3.24) is based on the technique of [85] and the following relation:

$$\begin{aligned} \int_{\mathbf{x}_k \neq \mathbf{0}_M} \|\mathbf{x}_k\|_2^2 p_{X|Y}(\mathbf{x}_k | \mathbf{Y}, \hat{\mathbf{x}}_{t,k}, \tau_t^2, \boldsymbol{\beta}_t) d\mathbf{x}_k &= \phi_{t,k}(\|\boldsymbol{\mu}_{t,k}\|_2^2 + \mathrm{tr}(\nu_{t,k} \mathbf{I}_M)) \\ &= \phi_{t,k}(\|\boldsymbol{\mu}_{t,k}\|_2^2 + M\nu_{t,k}). \end{aligned} \tag{3.25}$$

The initialization of the unknown parameters has an impact on the performance of the above EM update. The update is also affected by the fluctuation of the LSF coefficients owing to the distribution of potential users. Thus, we employ the following initialization strategy:

$$\beta_{0,k} = \frac{L}{N} \cdot \frac{\|\mathbf{x}_{\mathrm{MF},k}\|_2^2}{M}, \tag{3.26}$$

where $\mathbf{x}_{\mathrm{MF},k}$ denotes the k-th row of $\mathbf{X}_{\mathrm{MF}} \triangleq \mathbf{A}^{\mathrm{H}} \mathbf{Y}$.

Algorithm 3.1 EM-MMV-AMP

Input: Received signals $\mathbf{Y} \in \mathbb{C}^{L\times M}$, measurement matrix $\mathbf{A} \in \mathbb{C}^{L\times N}$, maximum number of iterations T_{amp}, and termination threshold η_{th}.

1: Initialize the iteration index t to 1, $\boldsymbol{\beta}_0$ as in (3.26), and the matrices as $\mathbf{X}^{(0)} = \mathbf{O}_{K\times M}$, $\mathbf{R}^{(0)} = \mathbf{Y}$.
2: **repeat**
3: Update the estimate $\mathbf{X}^{(t)}$ and the residual $\mathbf{R}^{(t)}$ using (3.7) and (3.8), respectively.
4: Obtain τ_t^2 based on (3.9).
5: Update the hyperparameters $\beta_{t,k}$ using (3.24), $\forall n$.
6: $t = t + 1$.
7: **until** $t \geq T_{\text{amp}}$ or $(*)$.
8: $(*)$ $\begin{cases} \|\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}\|_{\text{F}} / \|\mathbf{X}^{(t)}\|_{\text{F}} < \eta_{\text{th}} & \text{(EM-MMV-AMP)} \\ (\tau_t - \tau_{t-1})/\tau_{t-1} < \eta_{\text{th}} & \text{(MMV-AMP)} \end{cases}$

Output: The estimate of $\mathbf{X}$; $\hat{\mathbf{X}} = \mathbf{X}^{(t)}$, the estimated large-scale fading coefficients $\beta_{t,k}$, and $\hat{\mathbf{x}}_{t,k}$, $\forall n$.

In light of the above, Algorithm 3.1 summarizes the EM-MMV-AMP algorithm. Notice that, as shown in line 8 of Algorithm 3.1, EM-MMV-AMP exploits a termination criterion that is different from MMV-AMP.

3.3.3 Active User Detection by EM-MMV-AMP

In this subsection, we revisit and propose tailored decision rules to AUD exploiting the EM-MMV-AMP algorithm since EM-MMV-AMP estimates not only $\mathbf{X}$ in (3.34) but also the LSF coefficients β_k, unlike MMV-AMP.

As a conventional rule, we consider the following rule proposed in [54]:

$$\begin{cases} k \in \hat{\mathcal{A}}, & \text{if } \|\hat{\mathbf{x}}_{t,k}\|_2^2 \geq \theta_{t,k}, \\ k \notin \hat{\mathcal{A}}, & \text{otherwise}, \end{cases} \tag{3.27}$$

where $\hat{\mathcal{A}}$ denotes the set of estimated active users and

$$\theta_{t,k} = \left(\frac{1}{\tau_t^2} - \frac{1}{\beta_{t,k} + \tau_t^2}\right)^{-1} M \ln\left(1 + \frac{\beta_{t,k}}{\tau_t^2}\right). \tag{3.28}$$

As this rule mainly utilizes $\hat{\mathbf{x}}_{t,k}$ rather than $\beta_{t,k}$, we hence propose two decision rules making full use of the estimates of β_k.

Table 3.1 Simulation parameters

Number of potential users K	500, 1000
Number of active users K_{a}	100
Sequence length L	100
Transmitted power of each symbol ρ_{tx}	23 dBm
Power spectrum density of noise	-169 dBm/Hz
System bandwidth	1 MHz
Minimum distance between BS and users	0.05 km
Maximum distance between BS and users	1 km
Number of maximum iterations of the algorithms	200
Termination threshold of the algorithms	10^{-5}

The first rule only relies on τ_t^2:

$$\begin{cases} k \in \hat{\mathcal{A}}, & \text{if } \beta_{t,k} \geq 2\tau_t^2, \\ k \notin \hat{\mathcal{A}}, & \text{otherwise.} \end{cases} \tag{3.29}$$

The derivation of (3.29) is detailed in Appendix A. Note that this rule does not rely on the prior information of the channels.

The second rule can be utilized only when the BS knows the minimum value of β_k, *i.e.*, the LSF coefficient of the cell-edge user. This rule is given by:

$$\begin{cases} k \in \hat{\mathcal{A}}, & \text{if } \beta_{t,k} \geq \beta_{\min}, \\ k \notin \hat{\mathcal{A}}, & \text{otherwise,} \end{cases} \tag{3.30}$$

with $\beta_{\min}$ denoting the minimum value. This rule implies that the BS can employ the precise boundary and tends to avoid detecting as active ones.

3.3.4 Numerical Results

In this section, we evaluate the performance of the proposed scheme via computer simulations in terms of NMSE and the probabilities of MD and FA. The simulation parameters are listed in Table I, and the matrix $\mathbf{A}$ is obtained by normalizing each column comprising L randomly selected and reordered rows of an K-point DFT matrix

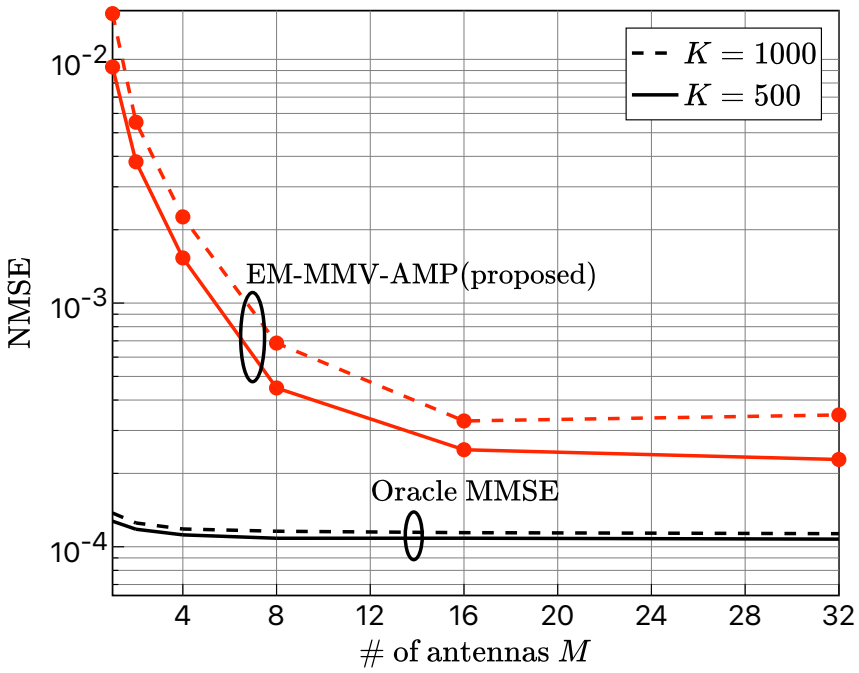

Fig. 3.1 NMSE performance for $K = 500$ and 1000.

for all simulations. The NMSE is defined as

$$\text{NMSE} \triangleq \mathbb{E}\left[\frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_{\text{F}}^2}{\|\mathbf{X}\|_{\text{F}}^2}\right]. \tag{3.31}$$

Moreover, MD is the event in which an active user is detected as a non-active user, whereas FA is the vice versa.

In this study, we consider both the MMV-AMP [54] and GMMV-AMP [64] as state-of-the-art schemes. For comparison, in GMMV-AMP, μ is fixed to zero and τ of each user differs, *i.e.*, $\tau \to \tau_{k,m}, \forall k, m$. In other words, the EM updates of GMMV-AMP are modified as follows:

$$\mu^{(t+1)} \leftarrow 0, \ \tau_{k,m}^{(t+1)} \leftarrow |A_{k,m}^{(t)}|^2 + B_{k,m}^{(t)}, \ \forall n, m, \tag{3.32}$$

where $A_{k,m}^{(t)}$ and $B_{k,m}^{(t)}$ are the quantities calculated using Eqs. (24) in [64]. Furthermore, the hyperparameters $\gamma_{k,m}, \forall n, m$ are initialized as the known activity ratio ϵ.

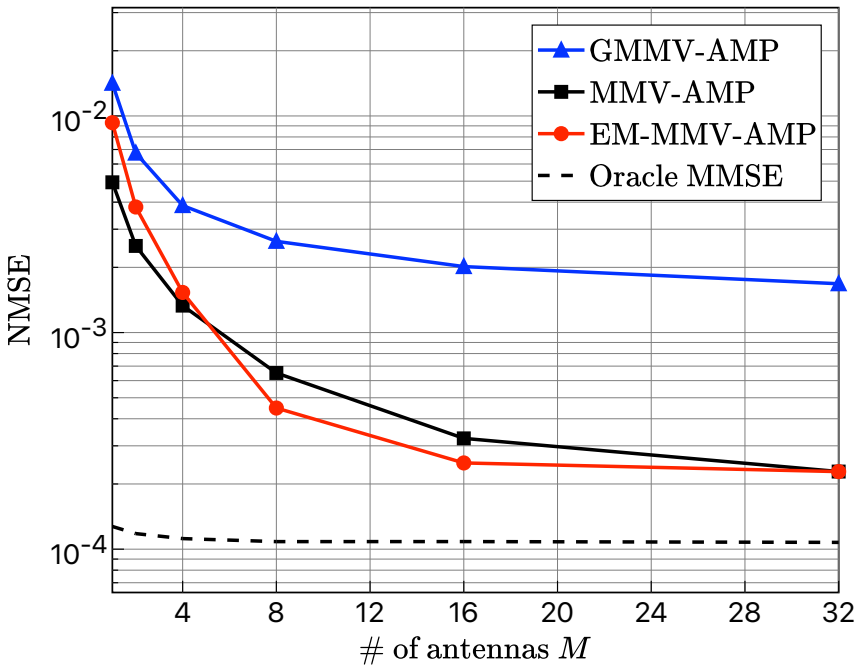

Fig. 3.2 NMSE performance of the proposed and conventional schemes.

NMSE Performance

Fig. 3.1 shows the NMSE performance of the proposed scheme when $K = 500$ and 1000. As a benchmark, this figure includes the performance of the MMSE estimation with the perfect knowledge of active users and LSF coefficients, denoted by "Oracle MMSE." This result indicates that the proposed initialization strategy (3.26) appropriately works for both $K = 500$ and 1000, enabling EM-MMV-AMP to approach the ideal performance as the number of antennas at the BS increases. Next, we investigate and discuss the performance focusing on the case of $K = 500$.

We compare the proposed and state-of-the-art schemes, as shown in Fig. 3.2). As shown in the figure, EM-MMV-AMP outperforms GMMV-AMP [64] owing to the full use of row sparsity in $\mathbf{X}$. Furthermore, the performance of the proposed scheme is comparable to that of MMV-AMP, exploiting the LSF coefficients of all users [54]. In particular, when $M \geq 8$, the proposed scheme slightly outperforms the conventional scheme because the rows of $\hat{\mathbf{X}}$ corresponding to non-active users tend to approach zero.

MD and FA Probabilities

To confirm the accuracy of AUD of the proposed scheme, we compared the decision rules considered in Section III-D. Fig. 3.3 shows the probabilities of MD and FA for

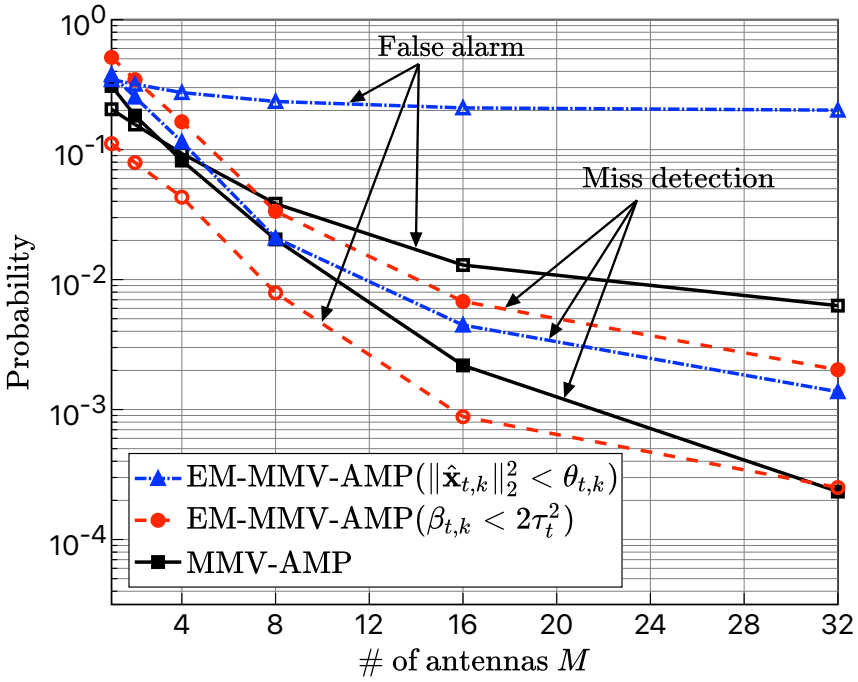

Fig. 3.3 MD and FA probabilities of the proposed scheme with (3.27) and (3.29).

(3.27) and (3.29) with the performance of the conventional MMV-AMP. According to Fig. 3.3, the FA probability of the proposed scheme with (3.27) remains above 10^{-1}, because both the estimates of $\mathbf{x}_k$ and β_k for a non-active user are close to zero. On the other hand, the decision rule of (3.29) lowers both the probabilities of MD and FA even though BS does not know β_k in advance. For instance, the proposed scheme can attain probabilities below 10^{-2} when the number of antennas is greater than 16. However, MMV-AMP requires at least 20 antennas at the BS to achieve FA probability below 10^{-2}.

Finally, we present the probabilities of MD and FA when the decision rule in (3.30) is adopted. The performance of the proposed scheme with AUD based on (3.30) is shown in Fig. 3.4. This result indicates that the decision rule of (3.30) can significantly reduce the FA probability because of the knowledge of the minimum value of β_k, whereas its MD probability is higher than 10^{-2} even when $M = 32$. In light of the above, we conclude that the best decision rule for AUD of EM-MMV-AMP is (3.29) among the three rules considered in this study.

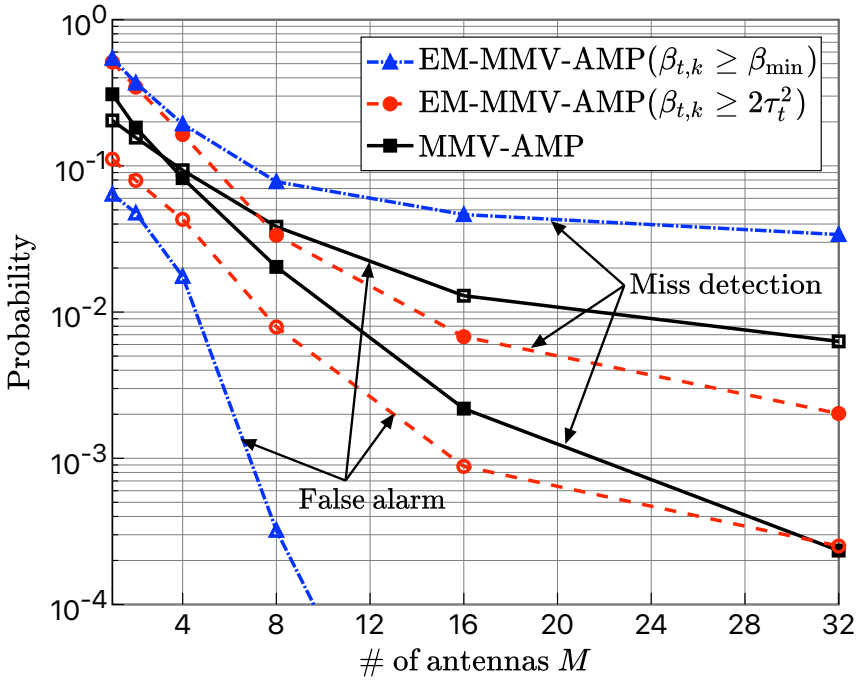

Fig. 3.4 MD and FA probabilities of the proposed scheme with (3.29) and (3.30).

3.4 Proposed AUD in the Presence of CFOs

In this section, we consider the AUD in the presence of CFOs under the assumption that the LSF coefficients of all users are ideally compensated, namely $\beta_k = 1$ for $k = 1, \ldots, K$.[1] Without loss of generality, we regard the angular frequency caused by the CFO ϖ_k as a random variable obeying a uniform distribution on the interval $[0, 2\pi)$ for $k = 1, \ldots, K$, which is considered to be the worst-case scenario. First, we transform the signal model (3.2) into the tailored one to address the AUD. Next, we formulate the problem of the AUD and propose the scheme to solve it efficiently, which exploits the CD method. Finally, we show superior performance of the proposed scheme via computer simulations.

3.4.1 Formulation Transformation

Let us focus on the diagonal matrix $\mathbf{A}_k = \mathrm{diag}(\mathbf{a}_k)$. Then, we take notice of the following relation [87]:

$$\mathbf{A}_k = \mathbf{F}_L \mathbf{C}_k \mathbf{F}_L^{\mathrm{H}}, \tag{3.33}$$

[1]The proposed AUD is applicable for the case that the LSF coefficients are not compensated. Thus, in Section 3.4.3, the impact of LSF on the performance of the proposed scheme will be confirmed.

where $\mathbf{C}_k \in \mathbb{C}^{L\times L}$ and $\mathbf{F}_L \in \mathbb{C}^{L\times L}$ denote the circulant matrix based on $\mathbf{a}_k$ and an $L\times L$ DFT matrix, whose the ℓ_2-norm of each column $\mathbf{f}_\ell \in \mathbb{C}^{L\times 1}$ is normalized, respectively.

Hereafter, we consider transforming the equation of received signals while dropping ξ from (3.2) for convenience, such as $\xi = 1$. Based on (3.2) and (3.33), the received signals multiplied by $\mathbf{F}_L^{\mathrm{H}}$ from the left-hand side can be expressed by

$$\begin{aligned}\underbrace{\mathbf{F}_L^{\mathrm{H}}\mathbf{Y}}_{\tilde{\mathbf{Y}}} &= \sum_{k\in\mathcal{A}} \mathbf{C}_k \underbrace{\mathbf{F}_L^{\mathrm{H}}\mathbf{b}_L(\varpi_k)\mathbf{h}_k^{\mathrm{T}}}_{\tilde{\mathbf{X}}_k} + \underbrace{\mathbf{F}_L^{\mathrm{H}}\mathbf{Z}}_{\tilde{\mathbf{Z}}} \\ &= [\mathbf{C}_1,\ldots,\mathbf{C}_K]\begin{bmatrix}\tilde{\mathbf{X}}_1\\ \vdots \\ \tilde{\mathbf{X}}_K\end{bmatrix} + \tilde{\mathbf{Z}} \\ &= \mathbf{C}\tilde{\mathbf{X}} + \tilde{\mathbf{Z}}, \end{aligned} \tag{3.34}$$

where $\tilde{\mathbf{Y}} = \mathbf{F}_L^{\mathrm{H}}\mathbf{Y}$, $\tilde{\mathbf{Z}} = \mathbf{F}_L^{\mathrm{H}}\mathbf{Z}$, $\mathbf{C} = [\mathbf{C}_1,\ldots,\mathbf{C}_K]$, and $\tilde{\mathbf{X}} = [\tilde{\mathbf{X}}_1^{\mathrm{T}},\ldots,\tilde{\mathbf{X}}_K^{\mathrm{T}}]$ with $\tilde{\mathbf{X}}_k = \mathbf{F}_L^{\mathrm{H}}\mathbf{b}_L(\varpi_k)\mathbf{h}_k^{\mathrm{T}}$. As $\mathbf{b}_L(\varpi_k)$ is a vector that has the same structure as a steering vector, $\mathbf{F}_L^{\mathrm{H}}\mathbf{b}_L(\varpi_k)$ for $k \in \mathcal{A}$ tends to have a certain single element that is relatively larger than the others [86].[2] Thus, the matrix $\tilde{\mathbf{X}}_k = \mathbf{F}_L^{\mathrm{H}}\mathbf{b}_L(\varpi_k)\mathbf{h}_k^{\mathrm{T}}$ for $k \in \mathcal{A}$ can be seen as a row-sparse matrix.

By introducing the variables $\gamma_k^\ell \in \mathbb{R}_+$ for $\ell = 1,\ldots,L$ and $k = 1,\ldots,K$, we rewrite $\tilde{\mathbf{X}}_k$ as follows

$$\begin{aligned}\tilde{\mathbf{X}}_k &= \mathbf{F}_L^{\mathrm{H}}\mathbf{b}_L(\varpi_k)\mathbf{h}_k^{\mathrm{T}} \\ &= \mathrm{diag}(\boldsymbol{\gamma}_k)^{\frac{1}{2}}\tilde{\mathbf{H}}_k, \end{aligned} \tag{3.35}$$

where $\boldsymbol{\gamma}_k \triangleq [\gamma_k^1,\ldots,\gamma_k^L]^{\mathrm{T}}$ and $\tilde{\mathbf{H}}_k \in \mathbb{C}^{L\times M}$ is the matrix whose each element obeys i.i.d. complex Gaussian distribution with zero mean and unit variance. Accordingly, $\tilde{\mathbf{Y}}$ can be written as

$$\begin{aligned}\tilde{\mathbf{Y}} &= \sum_{k\in\mathcal{A}} \mathbf{C}_k \mathrm{diag}(\boldsymbol{\gamma}_k)^{\frac{1}{2}}\tilde{\mathbf{H}}_k + \tilde{\mathbf{Z}} \\ &= \mathbf{C}\begin{bmatrix}\mathrm{diag}(\boldsymbol{\gamma}_1)^{\frac{1}{2}} & \mathbf{O}_{L\times L} & \cdots & \mathbf{O}_{L\times L} \\ \mathbf{O}_{L\times L} & \mathrm{diag}(\boldsymbol{\gamma}_2)^{\frac{1}{2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{O}_{L\times L} \\ \mathbf{O}_{L\times L} & \cdots & \mathbf{O}_{L\times L} & \mathrm{diag}(\boldsymbol{\gamma}_K)^{\frac{1}{2}}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{H}}_1\\ \vdots \\ \tilde{\mathbf{H}}_K\end{bmatrix} + \tilde{\mathbf{Z}} \end{aligned}$$

[2] This leads to a convex relaxation technique called *Lifting* [88, 89], which handles bilinear optimization problems by transforming them to a higher dimensional space. This technique is applied to the joint CFO and channel estimation problem in a single-user case [90, 91].

$$= \mathbf{C}\boldsymbol{\Gamma}^{\frac{1}{2}}\tilde{\mathbf{H}} + \tilde{\mathbf{Z}}, \tag{3.36}$$

where we have $\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1^{\mathrm{T}}, \ldots, \tilde{\mathbf{H}}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{KL\times M}$ and $\boldsymbol{\Gamma} \triangleq \mathrm{diag}(\boldsymbol{\gamma}) \in \mathbb{R}_+^{LK\times LK}$ with $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_{LK}]^{\mathrm{T}} = [\boldsymbol{\gamma}_1^{\mathrm{T}}, \ldots, \boldsymbol{\gamma}_k^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}_+^{LK\times 1}$. Notice that all elements of $\boldsymbol{\gamma}_k$ for $k \in \mathcal{A}$ are positive in (3.36), whereas only single element of $\boldsymbol{\gamma}_k$ for $k \in \mathcal{A}$ is positive in the problem considered in [51].

3.4.2 Proposed AUD

In a similar fashion to [51], we first formulate the estimation problem of $\boldsymbol{\Gamma}$ as *maximum likelihood* estimation problem. Let $\tilde{\mathbf{y}}_m \in \mathbb{C}^{L\times 1}$ denote the m-th column of $\tilde{\mathbf{Y}}$. Besides, each column of $\tilde{\mathbf{Y}}$ is assumed to obey i.i.d. multivariate complex Gaussian distribution:

$$\tilde{\mathbf{y}}_m \sim \mathcal{CN}\left(\mathbf{0}_L, \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_L\right), \tag{3.37}$$

where the covariance matrix can be computed by $\mathbb{E}[\tilde{\mathbf{y}}_m\tilde{\mathbf{y}}_m^{\mathrm{H}}]$ based on (3.36). For notational brevity, the covariance matrix in (3.37) is defined as $\boldsymbol{\Sigma} \triangleq \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_L$.

According to (3.37), the likelihood of $\tilde{\mathbf{Y}}$ given $\boldsymbol{\gamma}$ is

$$\begin{aligned} p(\tilde{\mathbf{Y}}|\boldsymbol{\gamma}) &= \prod_{m=1}^{M} \frac{1}{\pi\mathrm{det}(\boldsymbol{\Sigma})} \exp\left(-\tilde{\mathbf{y}}_m^{\mathrm{H}}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{y}}_m\right) \\ &= \frac{1}{(\pi\mathrm{det}(\boldsymbol{\Sigma}))^M} \exp\left(\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{H}}\right)\right). \end{aligned} \tag{3.38}$$

Hence, the ML estimation problem can be seen as the minimization problem of $-\ln p(\tilde{\mathbf{Y}}|\boldsymbol{\gamma})$.

In the straightforward approach [51], the aforementioned problem is equivalent to search the optimal $\boldsymbol{\gamma}$ in the space $[0, +\infty)^{LK\times 1}$. However, we can limit the feasible space of $\boldsymbol{\gamma}$ to $[0, L]^{LK\times 1}$ since the amplitude of elements of $\mathbf{F}_L^{\mathrm{H}}\mathbf{b}_L(\varpi_k)$ is bounded by $\sqrt{L}$. Notice that an inner product of column of $\mathbf{F}_L$ and $\mathbf{b}_L(\varpi_k)$ takes the maximum value $\sqrt{L}$ if and only if $\mathbf{b}_L(\varpi_k) = \sqrt{L}\mathbf{f}_\ell$. As a consequence, we can add the box-constraint and obtain the following optimization problem

$$\underset{\boldsymbol{\gamma}}{\text{minimize}} \;\; \ln\mathrm{det}(\boldsymbol{\Sigma}) + \frac{1}{M}\mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{H}}\right) \tag{3.39a}$$

$$\text{subject to} \;\; \gamma_i \in [0, L], \quad i = 1, \ldots, LK, \tag{3.39b}$$

$$\|\boldsymbol{\gamma}_k\|_0 \leq 1, \quad k = 1, \ldots, K. \tag{3.39c}$$

Algorithm 3.2 Coordinate descent to estimate $\boldsymbol{\gamma}$

Input: $\boldsymbol{\Sigma}_Y = \frac{1}{M}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^{\mathrm{H}}, \hat{\boldsymbol{\gamma}} = \mathbf{0}_{LK}, \hat{\boldsymbol{\Sigma}}^{-1} = \sigma_k^{-2}\mathbf{I}_{LK}$, The number of iterations T_{CD}.

1: **for** $i = 1, 2, \ldots, T_{\mathrm{CD}}$ **do**

2: Randomly select a permutation $i_1, i_2, \ldots, i_{LK}$ of the coordinate indices $\{1, 2, \ldots, LK\}$ of $\hat{\boldsymbol{\gamma}}$.

3: **for** $k = 1, 2, \ldots, LK$ **do**

4: $\delta_0 = \max\left\{\frac{\mathbf{c}_{i_k}^{\mathrm{H}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_Y\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{c}_{i_k} - \mathbf{c}_{i_k}^{\mathrm{H}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{c}_{i_k}}{(\mathbf{c}_{i_k}^{\mathrm{H}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{c}_{i_k})^2}, -\hat{\gamma}_{i_k}\right\}$

5: $\delta = \min\{\delta_0, L - \hat{\gamma}_{i_k}\}$

6: $\hat{\gamma}_{i_k} \leftarrow \hat{\gamma}_{i_k} + \delta$

7: $\hat{\boldsymbol{\Sigma}}^{-1} \leftarrow \hat{\boldsymbol{\Sigma}}^{-1} - \delta\frac{\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{c}_{i_k}\mathbf{c}_{i_k}^{\mathrm{H}}\hat{\boldsymbol{\Sigma}}^{-1}}{1+\delta\mathbf{c}_{i_k}^{\mathrm{H}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{c}_{i_k}}$

8: **end for**

9: **end for**

Output: $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_1, \ldots, \hat{\gamma}_{LK}]^{\mathrm{T}}$

This problem in (3.39) slightly differs from the one considered in [51] in that the parameters γ_i are bounded by the constraint (3.39b). In addition, the peak appearance in $\mathbf{F}_L^{\mathrm{H}}\mathbf{b}_L(\varpi_k)$ is reflected in the constraint of $\boldsymbol{\gamma}_k$ given in (3.39c). The effect of the constraint (3.39b) will be discussed in the following subsection.

To perform AUD at the BS effectively, we solve the optimization problem (3.39) by the CD method based on Algorithm 3.2, in a similar manner to [51]. Although this approach requires a matrix inversion, its computational complexity mainly comes from the matrix-vector multiplications in steps 4–7, whose complexity is $\mathcal{O}(L^2)$, thanks to the update based on the Sherman-Morrison rank-1 update identity [92]. Thus, the overall complexity is $\mathcal{O}(KL^3T_{\mathrm{CD}})$, where T_{CD} is the number of iterations. It is worth noting that although the complexity grows as the cubic of L, this algorithm is suitable to perform AUD in mMTC scenarios that K is much larger than L. The operation reflecting the box-constraint (3.39b) is step 5 in Algorithm 3.2. In the proposed AUD, the optimization problem (3.39) without the constraint (3.39c) is first solved by this algorithm, and the following decision step based on the estimate $\hat{\boldsymbol{\gamma}}$ is then performed to determine the estimated active user:

$$\hat{\mathcal{A}} = \left\{k \middle| \hat{\gamma}_k^{\max} \geq \eta\sigma_{\mathrm{n}}^2 \text{ and } \hat{\gamma}_k^{\max} = \max_{\ell=1,\ldots,L} \hat{\gamma}_{(k-1)L+\ell}\right\}, \tag{3.40}$$

where η is a predefined threshold. Unfortunately, a setting of the value of η in (3.40) needs to adjust its value by tracking the performance of the AUD for each setup numerically due to the performance tradeoff in AUD [50, 51].

3.4.3 Numerical Results

We investigate the performances of our proposal via computer simulations. For all simulations, the number of potential users K, the maximum number of iterations T_{CD} in Algorithm 3.2, and SNR are set to be 200, 10, and 10 dB, respectively. In addition, the first column vectors in the circulant matrices $\mathbf{C}_k$ are designed by the algorithm named *C-SIDCO* [93], and all column vectors are normalized. To evaluate the accuracy of AUD, the probabilities of MD and FA are respectively defined as

$$P_{MD} \triangleq \mathbb{E}\left[\frac{1}{K_{\mathrm{a}}}\left|\mathcal{A} \setminus \hat{\mathcal{A}}\right|\right], \tag{3.41}$$

$$P_{FA} \triangleq \mathbb{E}\left[\frac{1}{K - K_{\mathrm{a}}}\left|\hat{\mathcal{A}} \setminus \mathcal{A}\right|\right]. \tag{3.42}$$

Firstly, we confirm the performance degradation caused by CFO. The *receiver operating characteristics (ROCs)* of our proposal with $M = 16, 32$, and 64, where $L = 40$ and $K_{\mathrm{a}} = 20$, are shown in Fig. 3.5. As the benchmark with the absence of CFOs, the performance of the CD method proposed in [50] is also shown in the figure. It is obvious from Fig. 3.5 that the performance in the presence of CFOs is far from that of the benchmark even when the BS is equipped with more antennas, indicating that CFOs have a significant impact on the accuracy of AUD.

Fig. 3.6 shows the ROCs of our proposal and the conventional scheme based on non-negative least square (NNLS) [50] with $M = 16$, 32, and 64, where $L = 40$, $K_{\mathrm{a}} = 20$, and the value of η in (3.40) varies from 1 to 100. According to (3.40) and the figure, the probability of FA decreases, but that of MD increases as η increases. In addition, as is obvious from Fig. 3.6, the overall accuracy of AUD is gradually improved with the increase in M, and our proposal outperforms the conventional scheme. We next evaluate the performance gain thanks to the block-constraint (3.39b), where $K_{\mathrm{a}} > L$ and M is large, since it was shown that the suitable box-constraint enhances the performance of the CD approach in such a circumstance [52].

Fig. 3.7 illustrates the ROCs with $K_{\mathrm{a}} = 30, 40, 45$, and 50, where $L = 40$ and $M = 64$. As shown in the figure, the CD method with the constraint (3.39b) slightly outperforms the one without the constraint. This result implies that the performance improvement would be larger with the increase in K_{a}. On the contrary, the achievable accuracy of AUD degrades in such a case, and our proposal would be unlikely to function. Therefore, the results of the scheme without (3.39b) will be shown and discussed hereafter.

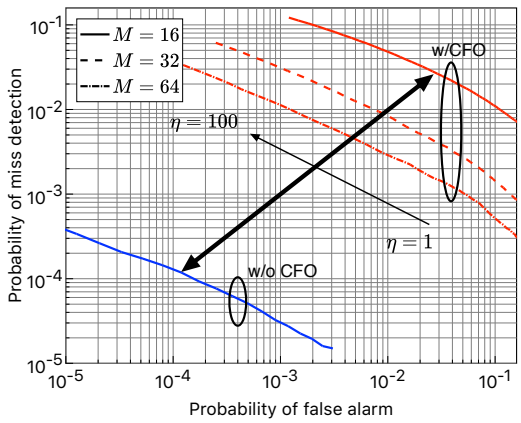

Fig. 3.5 ROCs of our proposal where $L = 40$ and $K_{\mathrm{a}} = 20$. The performance of the CD method [50] with $M = 16$ in the absence of CFOs is also shown.

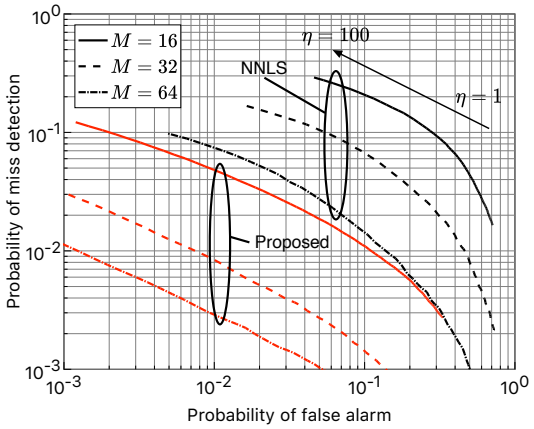

Fig. 3.6 ROCs of our proposal where $L = 40$ and $K_{\mathrm{a}} = 20$. The performances of the conventional scheme based on NNLS [50] are also shown.

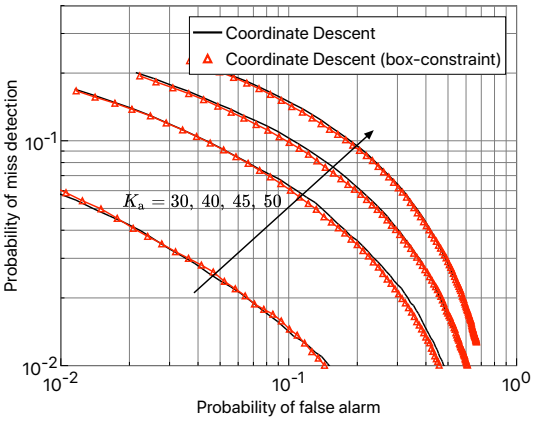

Fig. 3.7 Performance of our proposal where $L = 40$ and $M = 64$.

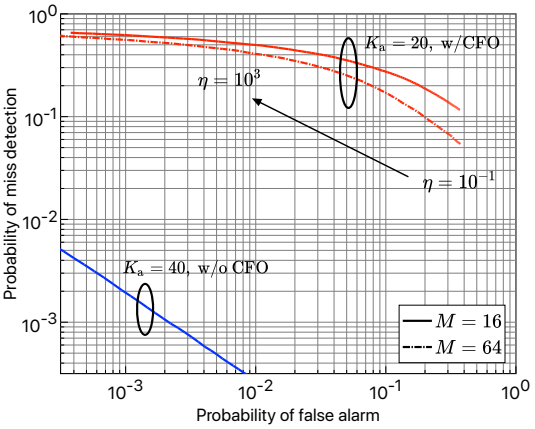

Fig. 3.8 Impact on LSF on the performance of our proposal where $L = 40$.

Here, we investigated the impact LSF on the performance of the proposed method. Fig. 3.8 shows the ROCs with $M = 16$ and 64, where $L = 40$ and $K_{\mathrm{a}} = 20$. Notice that the noise variance is determined following TABLE 2.2, and $\beta_k = -128.1 - 37.6 \log_{10}(d_k)[\mathrm{dB}]$, where d_k is uniformly distributed in $[0.2, 1]$ km. In addition, the performance of the CD method [50] with $M = 16$ and $K_{\mathrm{a}} = 40$ in the absence of CFOs is also shown. As seen from the figure, the performance in the presence of CFOs significantly degrades as compared to the case that $\beta_k = 1, \forall k$, while the benchmark can perform AUD accurately even when K_{a} is 40. This result implies that the coexistence of LSF and CFOs has a significant impact on the accuracy of AUD and that the AUD, which is suitable to such a case, is still a challenging task.

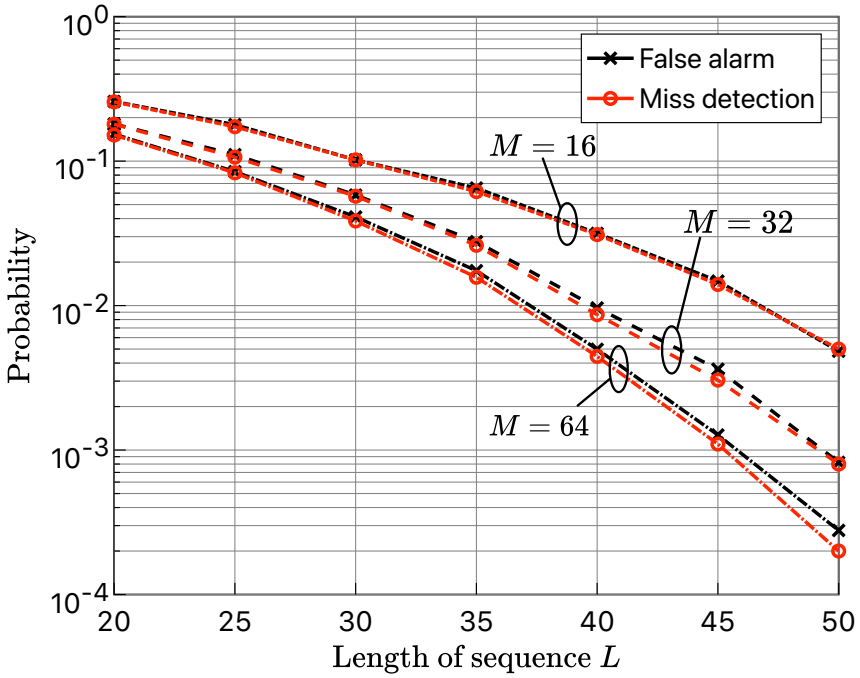

Fig. 3.9 Probabilities of MD versus L where $K_{\mathrm{a}} = 20$ and η takes the value satisfying $P_{MD} \approx P_{FA}$.

Moreover, we confirmed the effect of L on the accuracy of AUD. Fig. 3.9 demonstrates P_{MD} versus L, where $K_{\mathrm{a}} = 20$. The value of η is set to satisfy $P_{MD} \approx P_{FA}$ based on ROC for each setup. These simulation results indicate that the proposed scheme can reduce the overhead to accommodate a large number of users when M is large because the complexity does not depend on M. On the other hand, as shown in the figure, the performance improvement is not substantial, especially when M is large.

Finally, Fig. 3.10 shows the probability of MD when the number of active users K_{a} varies. The parameters L and M are fixed to 40 and 64, respectively. As obvious from the figure, our proposal achieves the low probability of MD when K_{a} is small, especially $K_{\mathrm{a}} \leq L/2$. In contrast, the increase of K_{a} results in the degradation of its performance, and the probability exceeds 10^{-1} when $K_{\mathrm{a}} > L$. This result suggests that GF-NOMA systems require longer spreading sequences to accommodate many active users efficiently if CFOs of users cannot be compensated for in advance.

3.5 Chapter Summary

In this chapter, we respectively examined GF-NOMA systems with spreading over the time-domain for two scenarios; 1) the BS does not know LSF coefficients of users

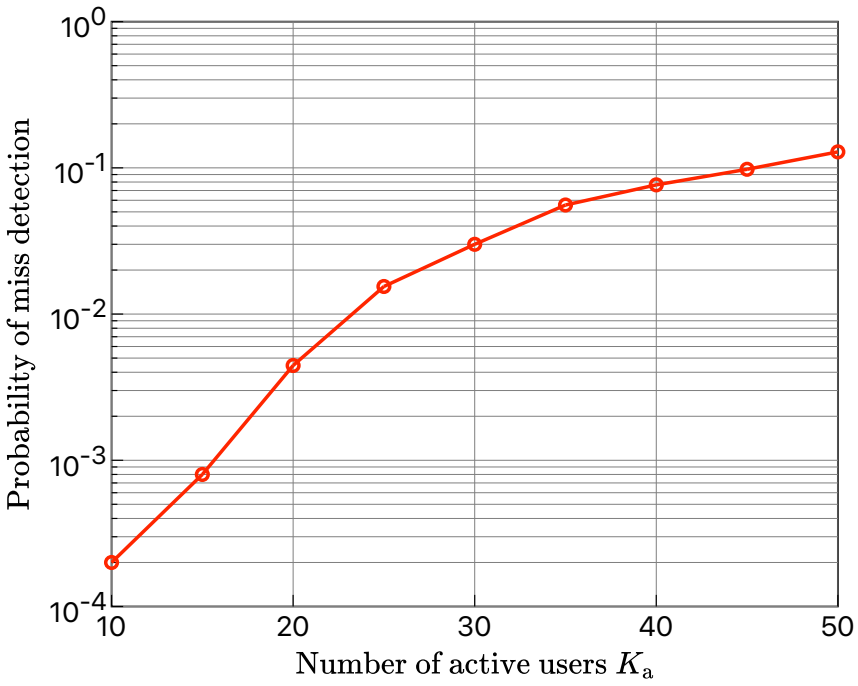

Fig. 3.10 Probability of MD versus K_{a} where $L = 40$, $M = 64$, and η takes the value satisfying $P_{MD} \approx P_{FA}$.

and 2) CFOs exist. We then proposed alternative schemes for each scenario and showed their superior performance via computer simulations. One of our proposals is the scheme to jointly perform AUD and CE via the EM-MMV-AMP algorithm that integrates the EM algorithm with the MMV-AMP algorithm. The other incorporates the transformation inspired by array-signal processing into the CD method, performing AUD in the presence of CFOs.

With the widespread use of IoT, future wireless communications systems are expected to be demanded to function without precise equipment, synchronization, and power control. Each proposed scheme exhibits higher estimation accuracy of active users as compared to conventional ones even when the systems cannot compensate for the effects of large-scale fading or CFOs. Thus, we conclude that this chapter contributes to a more practical receiver design for GF-NOMA systems with spreading over the time domain.

Chapter 4

Hyperparameter-Free Receiver for GF-NOMA Using Frequency Domain

Chapters 2 and 3 have investigated GF-NOMA systems with spreading over the *time* domain in the frequency synchronous and asynchronous scenarios. Since this transmission scheme requires a sufficiently long sequence in the time domain to accommodate massive users while keeping low AUD error, the systems need to widen the system bandwidth to meet the strict latency requirements. However, as well as many related works, the schemes considered in Chapters 2 and 3 are only applicable for a narrowband system under frequency-flat fading channels.

To this end, this chapter considers GF-NOMA systems exploiting an alternative transmission scheme to further reduce the latency. Concretely, we investigate GF-NOMA based on MIMO-OFDM with spreading over the *frequency* domain. Although several receivers for GF-NOMA with multicarrier transmission (MIMO-OFDM) have been considered, they would need a large number of antennas or prior knowledge of wireless channels. Motivated by this issue, *i.e.*, the fourth one in Section 1.6, we propose a hyperparameter-free receiver, which exploits the CD method utilizing the channel sparsity in the *delay* domain while avoiding any pre-tuning of hyperparameters and the need for prior knowledge of the noise and channels. It is revealed through computer simulations that the proposed receiver is superior to the classical algorithms using a block-sparsity or a sparsity of channels in the angular domain without prior knowledge of the noise and channels.

4.1 Overview of Related Works

As described in Section 1.5, many existing works, *e.g.*, [49–54, 63, 55–60], are designed for a narrowband system under frequency-flat fading channels. Although these approaches need to widen the system bandwidth to accommodate many users while keeping low latency, they are inapplicable for a wideband system under frequency-selective fading channels.

To overcome the issue, some existing works have considered an OFDM-based GF-NOMA system. CS-based random access with multicarrier transmission and its associated pilot design was investigated in [61], where the BS is assumed to be able to utilize the number of active users and the maximum delay spread among all users. However, this assumption is impractical because the BS cannot obtain these parameters in advance and utilize prior channel statistics as the reference, owing to sporadic traffic and mobility of uplink users. In addition, the authors of [62] have proposed the JACE based on MRAS recovery, which uses the inherent sparsity and low-rank structure of the millimeter-wave/Terahertz channels in the delay-angular domain, for millimeter-wave/Terahertz wideband GF-NOMA systems. Although this MRAS recovery can attain superior performance, it necessitates large numbers of antennas and the knowledge of the maximum number of propagation paths to exploit the low-rank structure.

On the other hand, the conventional receivers, *e.g.*, [55, 65], are applicable to OFDM-based GF-NOMA systems but they require an exhaustive search of their design parameters called *hyperparameters* for all (time-varying) system parameters such as SNR and user activity ratio. Doubtlessly, it would enforce a huge overhead to obtain those system parameters and tune the corresponding hyperparameters before transmission in practice. Hence, a *hyperparameter-free* receiver, which does not require tuning of any hyperparameters, is strongly desired.

Therefore, this chapter proposes a *hyperparameter-free* receiver for GF-NOMA with MIMO-OFDM, incorporating the pilot design considered in [94] to cope with issues described above. The main contributions can be summarized as follows:

- Our proposed receiver based on [95] utilizes the sparsity of channels in the delay domain and avoids any pre-tuning of resultant hyperparameters.
- We reveal that the notable mathematical connection between the methods of [95] and [50] and propose a hyperparameter-free receiver based on [50].

- We demonstrate that the proposed scheme is superior to the classical CS algorithms exploiting different sparseness and comparable to the state-of-the-art schemes for GF-NOMA systems.

4.2 System Model

In this chapter, we consider an uplink GF-NOMA system comprising K potential users, and a BS equipped with M antennas. Without loss of generality, the BS is supposed to be equipped with a one-dimensional uniform linear array (ULA) so that the antennas are separated by one-half wavelength. In addition, we assume that all users are time-synchronized and $K_{\mathrm{a}} \leq K$ users are active in each coherence time.

The uplink transmission is organized in OFDM symbols, where P pilot subcarriers are uniformly allocated to N subcarriers. All users share the same subcarrier locations for pilot transmission, and the subset of pilot subcarrier indices is denoted by $\mathcal{P} \subset \{1, 2, \ldots, N\}$, with $|\mathcal{P}| = P$. Let $\mathbf{Y} \in \mathbb{C}^{P \times M}$ denote the received signals at all the receiver antennas in the subset $\mathcal{P}$ after cyclic prefix (CP) removal and DFT modulation. Then, the received signals are given by

$$\begin{aligned} \mathbf{Y} &= \sum_{k \in \mathcal{A}} \mathrm{diag}(\mathbf{s}_k)\mathbf{G}_k + \mathbf{Z} \\ &\triangleq \sum_{k \in \mathcal{A}} \mathbf{S}_k \mathbf{G}_k + \mathbf{Z}, \end{aligned} \tag{4.1}$$

where $\mathcal{A}$ denotes the set of active users, $\mathbf{S}_k = \mathrm{diag}(\mathbf{s}_k)$ is a diagonal matrix based on the pilot sequence of user k, denoted by $\mathbf{s}_k \in \mathbb{C}^{P \times 1}$, which is assumed to be unimodular, *i.e.,* $|s_{k,p}| = 1$ for $p = 1, \ldots, P$. $\mathbf{G}_k = [\mathbf{g}_{k,1}, \ldots, \mathbf{g}_{k,P}]^{\mathrm{T}} \in \mathbb{C}^{P \times M}$ is the channel frequency response (CFR) between BS and user k over subcarriers $\mathcal{P}$. The matrix $\mathbf{Z} \in \mathbb{C}^{P \times M}$ represents the noise, whose all elements obey the i.i.d. complex Gaussian distribution with zero mean and variance σ_{n}^2. Here, we utilize the pilot design proposed in [94], assuming $K \leq P$ and $P < KN_{\mathrm{cp}}$, where N_{cp} denotes a CP length.

For the p-th pilot subcarrier ($p \in \mathcal{P}$), the sub-channel of the k-th user can be modeled as follows [96][1]

$$\mathbf{g}_{k,p} = \sum_{\ell=1}^{L_{\mathrm{path}}} \alpha_{k,\ell} \mathbf{b}_M(\varsigma_{k,\ell}) \mathrm{e}^{-j2\pi\tau_{k,\ell}\left(-\frac{B_{\mathrm{s}}}{2} + \frac{B_{\mathrm{s}}(pN/P-1)}{N}\right)} \in \mathbb{C}^{M \times 1}, \tag{4.2}$$

[1]Notice that the proposed scheme in Section 4.3 can be applied to the other channel models that does not need the BS to be equipped with ULA. To compare with the classical scheme that utilizes the channel sparsity in the angular domain, this chapter uses the channel model of [96].

where N/P is an integer, L_{path} represents the number of multi-path components (MPCs), $\alpha_{k,\ell} \sim \mathcal{CN}(0, 1/L_{\text{path}})$ and $\tau_{k,\ell} \in [0, N_{\text{cp}}/B_{\text{s}}]$ are the complex path gain and the path delay of the ℓ-th MPC, respectively. B_{s} is the two-sided bandwidth, and the antenna array response vector is denoted by $\mathbf{b}_M(\varsigma_{k,\ell}) = [1, \mathrm{e}^{-j2\pi\varsigma_{k,\ell}}, \ldots, \mathrm{e}^{-j2\pi(M-1)\varsigma_{k,\ell}}]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}$ with the phase difference between the received signal at adjacent antenna elements $\varsigma_{k,\ell}$.

The CFR can be represented by [97]

$$\mathbf{G}_k = \sqrt{N}\mathbf{F}_{P,N_{\text{cp}}}\mathbf{H}_k, \tag{4.3}$$

where $\mathbf{H}_k = [\mathbf{h}_k^1, \ldots, \mathbf{h}_k^M] \in \mathbb{C}^{N_{\text{cp}}\times M}$ denotes the channel impulse response (CIR) from the k-th user to the BS. The matrix $\mathbf{F}_{P,N_{\text{cp}}} \in \mathbb{C}^{P\times N_{\text{cp}}}$ is the sub-matrix of the $N \times N$ DFT matrix $\mathbf{F}_N$, and contains the P rows according to $\mathcal{P}$, and the first N_{cp} columns of $\mathbf{F}_N$.

Let $\bar{\mathbf{F}}_{P,N_{\text{cp}}} \in \mathbb{C}^{P\times N_{\text{cp}}}$ be the matrix $\mathbf{F}_{P,N_{\text{cp}}}$, with all column vectors normalized. Then, the received signals can be expressed with the CIRs as follows:

$$\begin{aligned}\mathbf{Y} &= \sum_{k\in\mathcal{A}} \mathbf{S}_k(\sqrt{P}\bar{\mathbf{F}}_{P,N_{\text{cp}}}\mathbf{H}_k) + \mathbf{Z} \\ &= \sum_{k\in\mathcal{A}} \mathbf{S}_k\bar{\mathbf{F}}_{P,N_{\text{cp}}}\mathbf{X}_k + \mathbf{Z} \in \mathbb{C}^{P\times M},\end{aligned} \tag{4.4}$$

where $\mathbf{X}_k \triangleq \sqrt{P}\mathbf{H}_k$ and $\sqrt{P}\bar{\mathbf{F}}_{P,N_{\text{cp}}} = \sqrt{N}\mathbf{F}_{P,N_{\text{cp}}}$. We define $\mathbf{A}_k = \mathbf{S}_k\bar{\mathbf{F}}_{P,N_{\text{cp}}} \in \mathbb{C}^{P\times N_{\text{cp}}}$, and thus, (4.4) can be simplified as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{Z}, \tag{4.5}$$

where $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_K]$ and $\mathbf{X} = [\mathbf{X}_1^{\mathrm{T}}, \ldots, \mathbf{X}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{KN_{\text{cp}}\times M}$.

Here, all column vectors of $\mathbf{A}$ are normalized. For massive MIMO systems, the inherent sparsity in the angular domain [64, 98] can be exploited. However, we take advantage of the row sparsity of $\mathbf{X}$ in (4.5) to avoid an increase in the number of required parameters in the estimation scheme.

4.3 Proposed Method

In this section, we propose two hyperparameter-free receivers. Unlike the existing approaches, including that of [61], our proposals exploit the row-sparsity instead of the block sparsity of $\mathbf{X}$ in (4.5). After formulating the problem for AUD and CE as

an ML problem [50], we reveal a nontrivial mathematical connection between ML [50] and SPARROW [95], which is the $\ell_{2,1}$ mixed-norm minimization problem.

4.3.1 ML-Based Problem Formulation

Based on [50], we define γ_k to represent the channel strengths involving the associated activity patterns, while $\mathbf{\Gamma} \triangleq \mathrm{diag}(\boldsymbol{\gamma})$ with $\boldsymbol{\gamma} \triangleq [\gamma_1, \ldots, \gamma_{KN_{\mathrm{cp}}}]^{\mathrm{T}}$. Then, (4.5) can be rewritten as

$$\mathbf{Y} = \mathbf{A}\mathbf{\Gamma}^{\frac{1}{2}}\bar{\mathbf{X}} + \mathbf{Z}, \tag{4.6}$$

where $\mathbf{X} = \mathbf{\Gamma}^{\frac{1}{2}}\bar{\mathbf{X}}$, and $\bar{\mathbf{X}}$ is the matrix whose each row obeys the complex standard Gaussian distribution. Based on the model given by (4.6), the ML estimation problem can be represented as [50, 51]

$$\underset{\mathbf{\Gamma}\in\mathbb{D}_+}{\text{minimize}} \ \ln\det(\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P) + \mathrm{Tr}((\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P)^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{Y}}), \tag{4.7}$$

where $\hat{\mathbf{\Sigma}}_{\mathbf{Y}} = \mathbf{Y}\mathbf{Y}^{\mathrm{H}}/M$ denotes the sample covariance matrix.

4.3.2 Nontrivial Connection Between ML and SPARROW

In this subsection, a mathematical connection between ML and SPARROW is presented, demonstrating how to obtain an estimate of $\mathbf{X}$ from (4.6). As the matrix $\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P$ is positive-definite, the following inequality holds

$$\ln\det(\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P) \leq \mathrm{Tr}(\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P - \mathbf{I}_P), \tag{4.8}$$

which readily yields the following relaxed problem of (4.7):

$$\underset{\mathbf{\Gamma}\in\mathbb{D}_+}{\text{minimize}} \ \mathrm{Tr}(\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}}) + \mathrm{Tr}((\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P)^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{Y}}). \tag{4.9}$$

Since the matrix $\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}}$ can be expressed as $\sum_{k=1}^{KN_{\mathrm{cp}}}\gamma_k\mathbf{a}_k\mathbf{a}_k^{\mathrm{H}}$, the first term of (4.9) can be transformed into $\sum_{k=1}^{KN_{\mathrm{cp}}}\gamma_k\|\mathbf{a}_k\|_2^2$ after some mathematical manipulations. Moreover, $\sum_{n=1}^{KN_{\mathrm{cp}}}\gamma_k\|\mathbf{a}_k\|_2^2$ is equal to $\sum_{k=1}^{KN_{\mathrm{cp}}}\gamma_k = \mathrm{Tr}(\mathbf{\Gamma})$ owing to the fact that all columns of the measurement matrix $\mathbf{A}$ are normalized, *i.e.*, $\|\mathbf{a}_k\|_2^2 = 1$. Interestingly, when the first term of (4.9) is expressed by the trace of $\mathbf{\Gamma}$ according to the above, (4.9)

can be regarded as a special case of SPARROW [95, Theorem 1] with $\lambda = \sigma_{\mathrm{n}}^2$:

$$\underset{\mathbf{\Gamma}\in\mathbb{D}_+}{\text{minimize}}\ \mathrm{Tr}(\mathbf{\Gamma}) + \mathrm{Tr}((\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \lambda\mathbf{I}_P)^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{Y}}), \tag{4.10}$$

which mathematically demonstrates that SPARROW is indeed a relaxed variate of ML.

According to [95], the estimate of $\mathbf{X}$ in (4.6) is given by

$$\hat{\mathbf{X}} = \hat{\mathbf{\Gamma}}\mathbf{A}^{\mathrm{H}}(\mathbf{A}\hat{\mathbf{\Gamma}}\mathbf{A}^{\mathrm{H}} + \lambda\mathbf{I}_P)^{-1}\mathbf{Y}, \tag{4.11}$$

where $\hat{\mathbf{\Gamma}}$ denotes an estimate of $\mathbf{\Gamma}$. Although SPARROW[95] and ML[50] can efficiently solve (4.7) and (4.10) by the CD method, they require a pre-determined parameter, such as the noise variance σ_{n}^2 or the pre-tuned hyperparameter λ, to execute. Hence, an approach without any knowledge of such parameters is needed, which will therefore be described in the following subsection.

4.3.3 Hyperparameter-Free Activity and Channel Estimation

To avoid any pre-determined parameters, the proposed scheme utilizes a reformulation of the matrix $\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \lambda\mathbf{I}_P$ with the second term denoting the covariance matrix of the additive white Gaussian noise (AWGN) noise at the receiver models the covariance matrix of each column of $\mathbf{Y}$, *i.e.,* $\mathbf{y}_m$. Let us assume that each diagonal entry of the covariance matrix of the noise can be replaced with an arbitrary variable, unlike in the original SPARROW formulation. Then, the covariance matrix of $\mathbf{y}_m$ can be modeled as

$$\begin{aligned}\mathbf{\Sigma} &= \mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_P\end{bmatrix} \\ &\triangleq \bar{\mathbf{A}}\bar{\mathbf{\Gamma}}\bar{\mathbf{A}}^{\mathrm{H}},\end{aligned} \tag{4.12}$$

where $\bar{\mathbf{A}} = [\mathbf{A}\ \mathbf{I}_P] \in \mathbb{C}^{P\times(KN_{\mathrm{cp}}+P)}$, and $\bar{\mathbf{\Gamma}}$ is the following diagonal matrix:

$$\bar{\mathbf{\Gamma}} = \begin{bmatrix} \mathbf{\Gamma} & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_P \end{bmatrix}. \tag{4.13}$$

Fortunately, even with this reformulation, $\mathbf{X}$ can be estimated from (4.11) as

$$\hat{\mathbf{X}} = \hat{\mathbf{\Gamma}}\mathbf{A}^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{Y}, \tag{4.14}$$

where $\hat{\mathbf{\Sigma}}$ denotes an estimate of the covariance matrix $\mathbf{\Sigma}$. It is worth noting that the above reformulation can be applied to the ML estimation since (4.7) involves a similarly structured covariance matrix, *i.e.,* $\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathrm{H}} + \sigma_{\mathrm{n}}^2\mathbf{I}_P$. Therefore, two different hyperparameter-free schemes according to ML and SPARROW can be presented for joint activity and channel estimation, respectively.

Following the CD method proposed in [95], the proposed scheme iteratively updates the variables in a coordinate fashion to minimize the objective function (4.7) or (4.10), where the covariance matrix including a hyperparameter is replaced with the matrix given in (4.12). In contrast to the approach in [95], the proposed CD method based on the parameter-free formulation described above jointly estimates the noise variance, activity patterns, and channel strengths, without pre-tuning of σ_{n}^2 or λ.

In light of all the above, an algorithmic flow of the proposed scheme is summarized in Algorithm 4.1. Notice that we assume that $\lambda_1 = \cdots = \lambda_P$ in the initialization step, which implicitly sets $\hat{\mathbf{\Sigma}}^{-1}$ to $\lambda_0^{-1}\mathbf{I}_P$ with $\lambda_0 = \|\mathbf{Y}\|_{\mathrm{F}}^2/(PM)$. In addition, in each outer iteration of the proposed algorithm, estimates of $\lambda_1, \ldots, \lambda_P$, *i.e.,* $\hat{\lambda}_1, \ldots, \hat{\lambda}_P$, are computed after the updates of the coordinates corresponding to $\boldsymbol{\gamma}$. The update of λ_p for $p = 1, 2, \ldots, P$ requires the p-th canonical basis vector with 1 only at its p-th entry and zero elsewhere, which is denoted by $\mathbf{e}_p$.

The computational complexity of the proposed method is mainly owing to the matrix-vector multiplications, and the complexity order required for each outer iteration is $\mathcal{O}((KN_{\mathrm{cp}} + P)P^2)$. While its complexity grows with the cubic of P, this can be acceptable for small P, and the proposed scheme can be considered efficient owing to the outstanding convergence rate of the CD method that was shown by numerical results in [95].

Algorithm 4.1 Hyperparameter-free CD method

Input: $\hat{\mathbf{\Sigma}}_{\mathbf{Y}} = \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}} \in \mathbb{C}^{P\times P}, \mathbf{A} \in \mathbb{C}^{P\times KN_{\mathrm{cp}}}, \lambda_0 = \frac{\|\mathbf{Y}\|_{\mathrm{F}}^2}{PM}$, The number of iterations T_{CD}.

Initialize $\hat{\mathbf{\Sigma}}^{-1} = \frac{1}{\lambda_0}\mathbf{I}_P, \hat{\boldsymbol{\gamma}} = \mathbf{0}_{KN_{\mathrm{cp}}}, \hat{\lambda}_1 = \cdots = \hat{\lambda}_P = \lambda_0$.
for $t = 1, 2, \ldots, T_{\mathrm{CD}}$ **do**
  Randomly select a permutation $i_1, i_2, \ldots, i_{KN_{\mathrm{cp}}}$ of the coordinate indices $\{1, 2, \ldots, KN_{\mathrm{cp}}\}$ of $\hat{\boldsymbol{\gamma}}$.
  for $k = 1, 2, \ldots, KN_{\mathrm{cp}}$ **do**
    $$\delta = \begin{cases} \max\left\{ \frac{\mathbf{a}_{i_k}^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{Y}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{a}_{i_k} - \mathbf{a}_{i_k}^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{a}_{i_k}}{(\mathbf{a}_{i_k}^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{a}_{i_k})^2}, -\hat{\gamma}_{i_k} \right\} & \text{(ML)} \\ \max\left\{ \frac{\sqrt{\mathbf{a}_{i_k}^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{Y}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{a}_{i_k}} - 1}{\mathbf{a}_{i_k}^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{a}_{i_k}}, -\hat{\gamma}_{i_k} \right\} & \text{(SPARROW)} \end{cases}$$
    $\hat{\gamma}_{i_k} \leftarrow \hat{\gamma}_{i_k} + \delta$
    $\hat{\mathbf{\Sigma}}^{-1} \leftarrow \hat{\mathbf{\Sigma}}^{-1} - \delta\frac{\hat{\mathbf{\Sigma}}^{-1}\mathbf{a}_{i_k}\mathbf{a}_{i_k}^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}}{1+\delta\mathbf{a}_{i_k}^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{a}_{i_k}}$
  end for
  for $p = 1, 2, \ldots, P$ **do**
    $$\delta = \begin{cases} \max\left\{ \frac{\mathbf{e}_p^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{Y}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{e}_p - \mathbf{e}_p^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{e}_p}{(\mathbf{e}_p^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{e}_p)^2}, -\hat{\lambda}_p \right\} & \text{(ML)} \\ \max\left\{ \frac{\sqrt{\mathbf{e}_p^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{\Sigma}}_{\mathbf{Y}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{e}_p} - 1}{\mathbf{e}_p^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{e}_p}, -\hat{\lambda}_p \right\} & \text{(SPARROW)} \end{cases}$$
    $\hat{\lambda}_p \leftarrow \hat{\lambda}_p + \delta$
    $\hat{\mathbf{\Sigma}}^{-1} \leftarrow \hat{\mathbf{\Sigma}}^{-1} - \delta\frac{\hat{\mathbf{\Sigma}}^{-1}\mathbf{e}_p\mathbf{e}_p^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}}{1+\delta\mathbf{e}_p^{\mathrm{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{e}_p}$
  end for
end for

Output: $\hat{\boldsymbol{\gamma}} \in \mathbb{R}^{KN_{\mathrm{cp}}\times 1}, \hat{\mathbf{\Sigma}}^{-1} \in \mathbb{C}^{P\times P}$.

After processing the CD method, we obtain an estimate of $\mathbf{X}$ by (4.14). Furthermore, we determine the estimated set of $\mathcal{A}$ as follows

$$\hat{\mathcal{A}} = \left\{ k \mid \|\hat{\mathbf{H}}_k\|_{\mathrm{F}} \geq \eta^* \|\hat{\mathbf{H}}_{\max}\|_{\mathrm{F}} \text{ and } \hat{\mathbf{H}}_{\max} = \max_{k=1,\ldots,K} \|\hat{\mathbf{H}}_k\|_{\mathrm{F}} \right\}, \tag{4.15}$$

with $\eta^* = 0.1$ denoting the ratio of the minimum and maximum Frobenius norms of the channel coefficients[2].

[2]Although η^* is a tunable parameter, the search for an optimal value is beyond the scope of this study. If this search is conducted, a ROC needs to be evaluated [50].

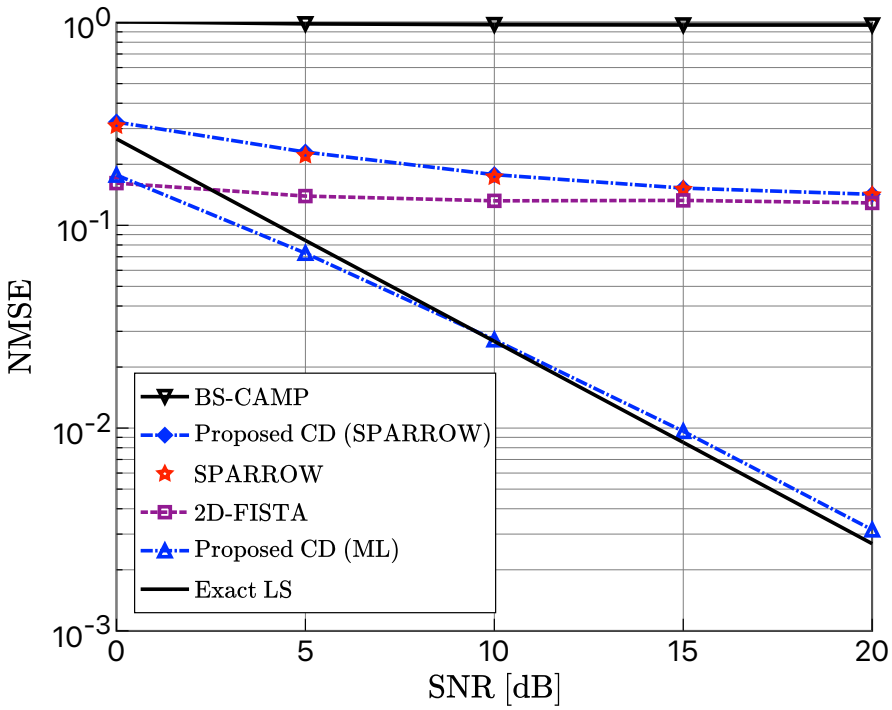

Fig. 4.1 NMSE performance of the proposed scheme and classical CS algorithms for $P = 64$.

4.4 Numerical Results

In this section, we investigate the NMSE performance and the activity error rate (AER), which is the probability of the activity pattern of each user to be detected incorrectly. In this chapter, we define the NMSE as

$$\mathrm{NMSE} \triangleq \mathbb{E}\left[\frac{\|\mathbf{H}-\hat{\mathbf{H}}\|_{\mathrm{F}}^2}{\|\mathbf{H}\|_{\mathrm{F}}^2}\right]. \tag{4.16}$$

For the simulation results, the bandwidth is $B_{\mathrm{s}} = 10$ MHz, and the number of significant paths L_{path} is 6. The system adopts OFDM, where $N = 1024$ subcarriers and a CP of length $N_{\mathrm{cp}} = 32$ are employed. The SNR is defined by the ratio of the pilot norm to the noise variance as $\mathrm{SNR} \triangleq \|\mathbf{s}_k\|_2^2/(P\sigma_{\mathrm{n}}^2) = 1/\sigma_{\mathrm{n}}^2$. We assume that the BS is equipped with a one-dimension ULA with $M = 32$ elements with half-wavelength spacing, and $K = 64$ potential users exist while $K_{\mathrm{a}} = 8$ users are active unless otherwise specified. In addition, $\varsigma_{k,\ell} = \sin(\varphi_{k,\ell})/2$ with the angle of arrival of the k-th user's ℓ-th MPC $\varphi_{k,\ell}$, and $\varphi_{k,\ell}$ is supposed to be uniformly distributed in $[-\pi, \pi]$. The number of iterations of the proposed scheme, denoted by T_{CD}, is set to 10.

Fig. 4.1 shows the NMSE performance for $P = 64$. As a benchmark, we evaluate the performance of a complex approximate message passing (CAMP) exploiting block-sparsity, namely, BS-CAMP, with 50 iterations. This scheme does not require prior impractical knowledge at the BS, unlike the approach in [61]. Moreover, we show the performance of the scheme of [99], *i.e.*, 2D-FISTA, with 100 iterations to evaluate the case that the sparsity in the angular domain is exploited in the estimation[3]. The performance of the least squares (LS) estimator, where the BS knows the indices of non-zero rows of $\mathbf{H}$, and that of the CD method of [95], with pre-tuned λ and 10 iterations, are denoted by Exact LS and SPARROW, respectively. In comparison, the proposed CD method with update rule of [95], denoted by Proposed CD (SPARROW), can outperform BS-CAMP, and its performance can approach that of SPARROW, whereas the one using update rule of [50], denoted by Proposed CD (ML), is significantly superior to BS-CAMP and 2D-FISTA in terms of performance. Our results thus imply that the $\ell_{2,1}$ mixed-norm minimization is a more suitable approach for GF-NOMA systems with MIMO-OFDM, compared to classical schemes utilizing block-sparsity and 2D-CS. Furthermore, while our proposed method does not require hyperparameters, its performance approaches that of Exact LS.

Moreover, the NMSE performance of the proposed and conventional schemes for $P = 64$ is shown in Fig. 4.2. As conventional schemes, we evaluate the performance of the schemes proposed in [50], *i.e.*, NNLS and ML, where the number of iterations is set to 10. These results show that the proposed scheme can outperform NNLS significantly and is comparable to ML. Note that our objective is to achieve superior performance without any hyperparameters that need to be designed via an exhaustive search to yield a low estimation error. In light of the above, our proposed receiver can meet such an objective and is comparable to the state-of-the-art receiver for GF-NOMA systems. Next, we demonstrate the accuracy of activity detection of the proposed scheme.

Fig. 4.3 represents the AER versus the number of active users K_{a}, where $P = 64$ and SNR $= 5$ dB. As the activity detection depends on the accuracy of channel estimation, we focus on our proposal with update rule of [50] and ML. As seen from the figure, the accuracy of activity detection degrades as the number of active users increases. However, the AER of the proposed method can reach approximately 10^{-2}, even when 20% users are active, *i.e.,* $K_{\mathrm{a}}/K = 13/64 \approx 0.2$. Although the conventional studies on GF-NOMA, *e.g.*, [54, 65], consider the case where 10% of users or fewer are active,

[3]Although the Lipschitz constant is calculated by $\max_{\|\mathbf{X}\|_{\mathrm{F}}=1} \|\mathbf{A}^{\mathrm{H}}\mathbf{A}\mathbf{X}\|_{\mathrm{F}}$ in [99], it is computed via $\|\mathbf{A}^{\mathrm{H}}\mathbf{A}\|_{\mathrm{F}}$ based on the relation $\|\mathbf{A}^{\mathrm{H}}\mathbf{A}\mathbf{X}\|_{\mathrm{F}} \leq \|\mathbf{A}^{\mathrm{H}}\mathbf{A}\|_{\mathrm{F}}\|\mathbf{X}\|_{\mathrm{F}}$ since $\|\mathbf{X}\|_{\mathrm{F}} = 1$ is not usually satisfied in this study.

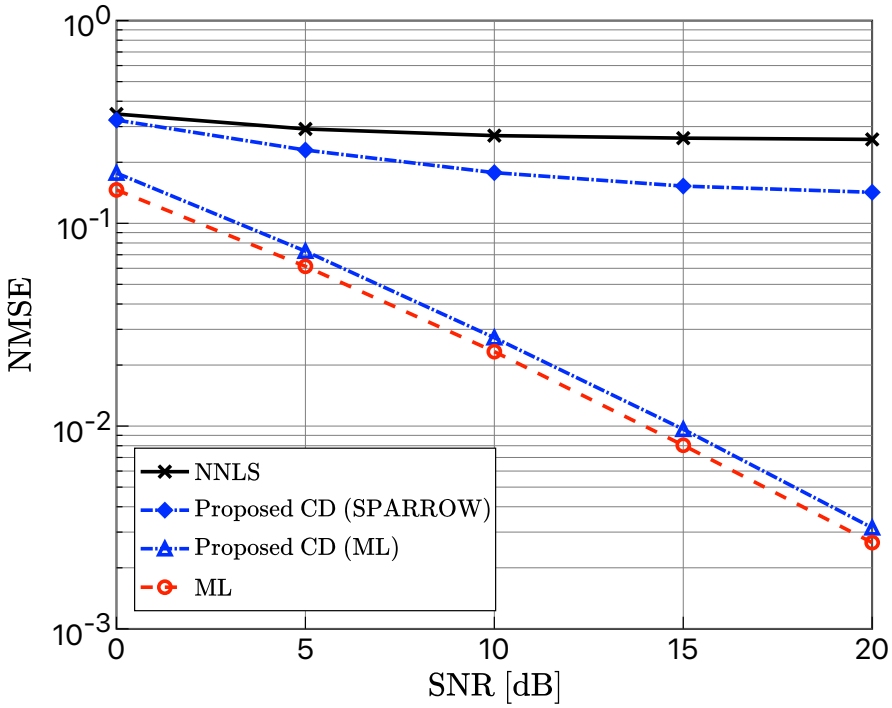

Fig. 4.2 NMSE performance of the proposed and conventional schemes for $P = 64$.

the proposed scheme can achieve low error rate even under tougher conditions. This indicates that our proposed scheme is robust against the variation of the traffic.

4.5 Chapter Summary

This chapter investigated GF-NOMA systems employing the spreading pilot sequence over the frequency domain, unlike the previous chapters, to further reduce access latency. We proposed a hyperparameter-free receiver, which takes advantage of the channel sparsity in the delay domain and avoids pre-tuning of parameters for accurate estimation. It is confirmed that the proposed scheme outperforms the conventional algorithms utilizing a block-sparsity or a sparsity of channels in the angular domain via computer simulations. Therefore, this chapter contributes to the reduction of access latency of GF-NOMA systems while taking into account the feasibility.

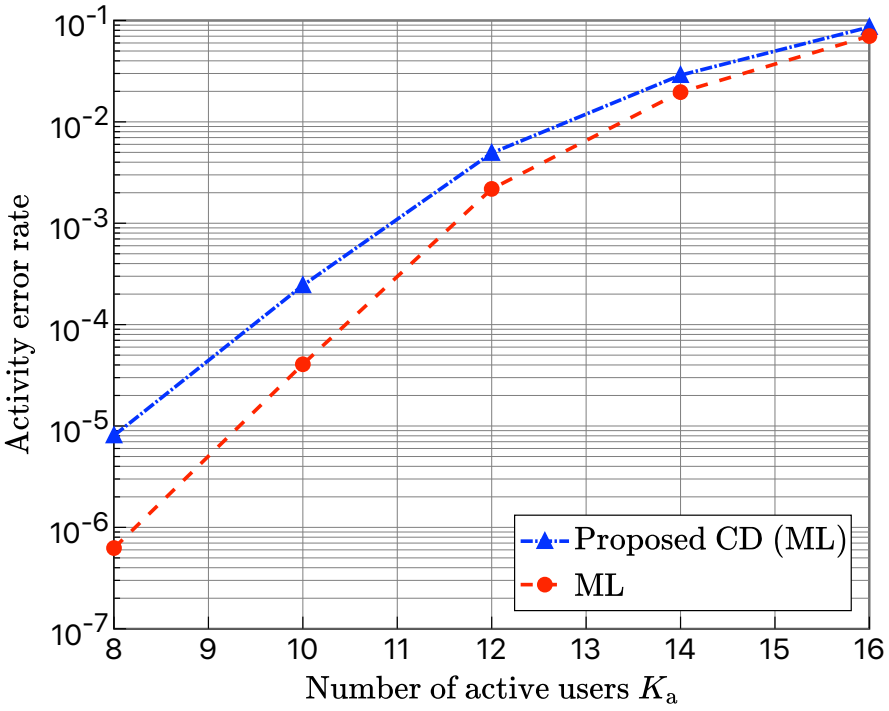

Fig. 4.3 AER of the proposed scheme for $P = 64$ and SNR $= 5$ dB.

Chapter 5

Massive GF-NOMA Using Time and Frequency Domains

In previous chapters, we have focused on the GF-NOMA systems with spreading over the time or frequency domain. However, each of the two approaches has an inherent issue when the systems accommodate many users efficiently. It is difficult for the approach using the time domain only to meet stringent latency requirements since the system has to use a long sequence. On the other hand, in the approach using the frequency domain only, while latency requirements are easily satisfied, the number of users that the system can accommodate is limited. This limitation comes from the use of only one OFDM symbol.

To overcome these issues, this chapter investigates GF-NOMA systems that make full use of both the time and frequency domains. Although a promising OFDM-based GF-NOMA scheme has been proposed in [64], it still requires a large number of antennas at the BS and dozens of OFDM symbols perform accurate estimation in supporting large numbers of users. We thus introduce a tailored signal model for the JACE that exploits the channel sparsity in the delay domain in the same fashion as Chapter 4. Moreover, we propose a system design that takes advantage of the signal model in the delay domain and theoretical analyses of the family of AMP algorithms so as to perform reliable JACE while reducing access latency. Accordingly, we consider an efficient MUD that exploits the corresponding channel responses reconstructed from the estimates in the JACE. Numerical results have demonstrated that the proposed scheme can outperform state-of-the-art approach in handling low-latency massive-access scenarios. Finally, we discuss the advantageous region of each of the three spreading patterns considered in this dissertation.

5.1 Background

As introduced in Section 1.1, to support time-sensitive applications such as motion control, future wireless systems will need to satisfy the heterogeneous requirements, especially the combination of mMTC and URLLC, namely massive URLLC [5–7]. To this end, GF-NOMA has been attracted as a promising scheme to meet massive connectivity and low latency and has been actively investigated in the literature. However, none of the related works described in Section 1.5 have addressed the concrete system design to satisfy the latency requirements, whereas such a design is essential to realize massive low-latency communications.

Therefore, this chapter considers a design of the GF-NOMA system that makes full use of both the time and frequency domains to satisfy the heterogeneous requirements of massive connectivity and low latency. The main contributions of this chapter are summarized as follows:

- **GF-NOMA with time-and-frequency spreading:** Our proposed GF-NOMA makes full use of radio resources in both the time and frequency domains to accommodate large numbers of users while reducing the access latency. To this end, we introduce a tailored signal model for the JACE that exploits the channel sparsity in the *delay* domain caused by limitations on the number of significant channel taps [100].

- **Data transmission for achieving moderate data-rate:** In conventional schemes, *e.g.*, [65, 67, 69], data symbols are spread over the time domain in the same manner as in the pilot, significantly lowering the data rate. To improve the data rate, the proposed scheme applies a spreading pattern that differs between the pilot and transmitted data.

- **Feasible sequence and system designs for low-latency massive access:** In a departure from the DCS theory-based design in [64], we propose a sequence design that takes advantage of the signal model in the delay domain. We further propose a tailored GF-NOMA design that performs reliable JACE while reducing access latency. The proposed design is based on an asymptotic analysis of a sparse recovery technique, namely *phase transition*.

- **Efficient JACE and MUD:** To manage the large-CS problem, we utilize an approach based on the GMMV-AMP algorithm [64] that can perform AUD and CE efficiently without a priori information on the channels. We further propose an efficient MUD via Gaussian belief propagation (GaBP) [101, 102], which

exploits the corresponding channel responses reconstructed from the estimates in the JACE.

- **Discussion of the design of GF-NOMA:** In this dissertation, we have considered GF-NOMA systems that employ three different spreading patterns: 1) over the time domain only, 2) over the frequency domain only, and 3) over both the time and frequency domains. Hence, we show the advantageous region of each of the three spreading patterns.

5.2 System Model

We consider an uplink GF-NOMA system comprising K single-antenna potential users and a common BS equipped with an M-antenna ULA, whose elements are equally spaced at a half wavelength. The uplink transmission is organized into OFDM symbols with N subcarriers (samples) and a subcarrier spacing of ΔB. All users use resources following the common frame structure illustrated in Fig. 5.1, in which $P \leq N$ subcarriers are uniformly allocated as pilot subcarriers and the others are used for data transmission. We assume that the subcarriers for the data component are divided into multiple groups, each of which comprises D subcarriers, and focus on discussing a single group of the data component throughout this chapter. The subsets of the indices of the pilot and data subcarriers are defined as $\mathcal{P}$ and $\mathcal{D}$, respectively.

The uplink transmission format conforms to the radio frame structure shown in Fig. 5.2, in which each frame comprises 10 subframes, each subframe comprises $U = 2^u (u = 0, 1, \ldots)$ time slots, and each time slot comprises 14 OFDM symbols. We consider the transmission within a single subframe that comprises $T = U \cdot 14$ OFDM symbols with a subcarrier spacing of $\Delta B = U \cdot 15$ kHz.

5.2.1 Signal Model in the Pilot

Let $\mathbf{Y}_{\mathrm{p}}^{(t)} \in \mathbb{C}^{P \times M}$ denote the received signals in subset $\mathcal{P}$ at the t-th OFDM symbol after CP removal and DFT modulation. The received signals are then given by

$$\begin{aligned} \mathbf{Y}_{\mathrm{p}}^{(t)} &= \sum_{k \in \mathcal{A}} \mathrm{diag}(\mathbf{s}_k^{(t)}) \mathbf{G}_k + \mathbf{Z}_{\mathrm{p}}^{(t)} \\ &= \sum_{k \in \mathcal{A}} \mathbf{S}_k^{(t)} \mathbf{G}_k + \mathbf{Z}_{\mathrm{p}}^{(t)}, \end{aligned} \tag{5.1}$$

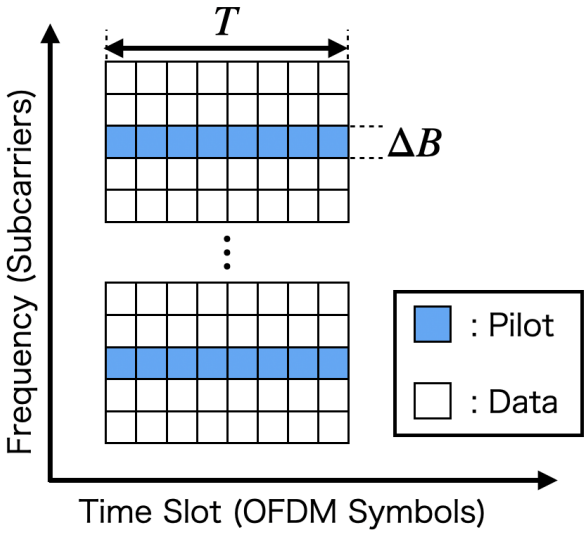

Fig. 5.1 Illustration of the uplink signal model.

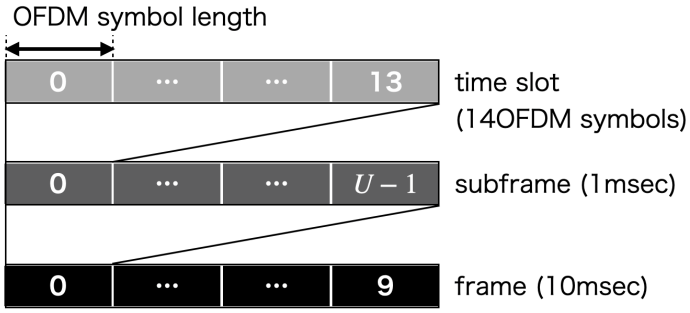

Fig. 5.2 Radio frame structure in 5G NR [103].

where $\mathcal{A}$ denotes the set of active users and $\mathbf{S}_k^{(t)} = \mathrm{diag}(\mathbf{s}_k^{(t)})$ is a diagonal matrix based on the pilot sequence of user k, denoted by $\mathbf{s}_k^{(t)} \in \mathbb{C}^{P\times 1}$. In this chapter, we assume that $\mathbf{s}_k^{(t)}$ is unimodular, satisfying $\|\mathbf{s}_k^{(t)}\|_2 = \sqrt{P}$. In addition, $\mathbf{G}_k = [\mathbf{g}_{k,1}, \ldots, \mathbf{g}_{k,P}]^{\mathrm{T}} \in \mathbb{C}^{P\times M}$ is the CFR between the BS and user k, and the matrix $\mathbf{Z}_{\mathrm{p}}^{(t)} \in \mathbb{C}^{P\times M}$ represents the noise, in which the elements obey a complex Gaussian distribution with zero mean and variance σ_{n}^2.

We define $\mathbf{H}_k \in \mathbb{C}^{L\times M}$ as the CIR from the k-th user to the BS with $L = \lceil \tau_{\max} B_{\mathrm{s}} \rceil + 1$ taps, where B_{s} and $\tau_{\max}$ are the system bandwidth and maximum path delay, respectively. Then, the CFR can be expressed in terms of the CIR as follows:

$$\begin{aligned} \mathbf{G}_k &= \sqrt{N}\mathbf{F}_{P,L}\mathbf{H}_k \\ &= \sqrt{P}\bar{\mathbf{F}}_{P,L}\mathbf{H}_k, \end{aligned} \tag{5.2}$$

where $\mathbf{F}_{P,L} \in \mathbb{C}^{P\times L}$ is a matrix containing P rows according to $\mathcal{P}$ and the first L columns of the $N \times N$ DFT matrix $\mathbf{F}_N$ and $\bar{\mathbf{F}}_{P,L}$ is a column-normalized version of $\mathbf{F}_{P,L}$. For the sake of simplicity, we assume that $\tau_{\max}$ is smaller than a CP duration.

Following [96], the sub-channel of the k-th user for the p-th pilot subcarrier can be modeled as

$$\mathbf{g}_{k,p} = \sum_{\ell=1}^{L_{\text{path}}} \alpha_{k,\ell}\mathbf{b}_M(\varsigma_{k,\ell})\mathrm{e}^{-j2\pi\tau_{k,\ell}\left(-\frac{B_\text{s}}{2}+\frac{B_\text{s}(p-1)}{N}\right)} \in \mathbb{C}^{M\times 1}, \tag{5.3}$$

where L_{path}, $\alpha_{k,\ell} \sim \mathcal{CN}(0, 1/L_{\text{path}})$, and $\tau_{k,\ell}$ are the number of MPCs, the complex path gain, and the path delay of the ℓ-th MPC, respectively. B_s is the two-sided bandwidth, and the antenna array response vector is denoted by $\mathbf{b}_M(\varsigma_{k,\ell}) = [1, \mathrm{e}^{-j2\pi\varsigma_{k,\ell}}, \ldots, \mathrm{e}^{-j2\pi(M-1)\varsigma_{k,\ell}}]^\text{T} \in \mathbb{C}^{M\times 1}$ with the phase difference between the received signal at adjacent antenna elements $\varsigma_{k,\ell}$. Under (5.3) and because the number of significant paths in the delay domain is limited [100], the CIR $\mathbf{H}_k$ is row-sparse and the number of non-zero rows in $\mathbf{H}_k$ is less than L_{path}.

Given the specifications above, (5.1) can be rewritten as follows:

$$\begin{aligned}\mathbf{Y}_\text{p}^{(t)} &= \sum_{k\in\mathcal{A}} \mathbf{S}_k^{(t)}\bar{\mathbf{F}}_{P,L} \times \sqrt{P}\mathbf{H}_k + \mathbf{Z}_\text{p}^{(t)} \\ &= \sum_{k\in\mathcal{A}} \mathbf{A}_k^{(t)}\tilde{\mathbf{H}}_k + \mathbf{Z}_\text{p}^{(t)} \\ &= \mathbf{A}^{(t)}\tilde{\mathbf{H}} + \mathbf{Z}_\text{p}^{(t)},\end{aligned} \tag{5.4}$$

where $\mathbf{A}_k^{(t)} \triangleq \mathbf{S}_k^{(t)}\bar{\mathbf{F}}_{P,L}$, $\tilde{\mathbf{H}}_k \triangleq \sqrt{P}\mathbf{H}_k$, $\mathbf{A}^{(t)} = [\mathbf{A}_1^{(t)}, \ldots, \mathbf{A}_K^{(t)}] \in \mathbb{C}^{P\times KL}$, and $\tilde{\mathbf{H}} = [\tilde{\mathbf{H}}_1^\text{T}, \ldots, \tilde{\mathbf{H}}_K^\text{T}]^\text{T} \in \mathbb{C}^{KL\times M}$. Furthermore, based on the assumption that $\|\mathbf{s}_k^{(t)}\|_2 = \sqrt{P}$ and each column of $\bar{\mathbf{F}}_{P,L}$ is normalized, $\mathbf{A}^{(t)}$ is a unit-norm matrix.

To make use of the time domain, we consider the following signal model representing the signals received through T OFDM symbols:

$$\begin{aligned}\mathbf{Y}_\text{p} &= [(\mathbf{Y}_\text{p}^{(1)})^\text{T}, \ldots, (\mathbf{Y}_\text{p}^{(T)})^\text{T}]^\text{T} \\ &= [(\mathbf{A}^{(1)})^\text{T}, \ldots, (\mathbf{A}^{(T)})^\text{T}]^\text{T}\tilde{\mathbf{H}} + [(\mathbf{Z}_\text{p}^{(1)})^\text{T}, \ldots, (\mathbf{Z}_\text{p}^{(T)})^\text{T}]^\text{T} \\ &= \tilde{\mathbf{A}}\tilde{\mathbf{H}} + \mathbf{Z}_\text{p} \\ &= \mathbf{A}\mathbf{X} + \mathbf{Z}_\text{p} \in \mathbb{C}^{PT\times M},\end{aligned} \tag{5.5}$$

where the matrix $\tilde{\mathbf{H}} \in \mathbb{C}^{KL\times M}$ is equivalent to that in (5.4), $\mathbf{A} = \tilde{\mathbf{A}}/\sqrt{T} \in \mathbb{C}^{PT\times KL}$, and $\mathbf{X} = \sqrt{T}\tilde{\mathbf{H}}$. Then, all columns of $\mathbf{A}$ are normalized. Note that the model presented

in (5.5) implies that the proposed GF-NOMA system employs different sequences across different time slots (equivalently, OFDM symbols) to enlarge the dimension of the measurement—*i.e.*, the number of rows in $\mathbf{Y}_{\mathrm{p}}$—thereby enabling reliable CS-based sparse recovery.

5.2.2 Signal Model in the Data

The received signal corresponding to the data component at the t-th OFDM symbol and the m-th receiving antenna is given by

$$\begin{aligned}\mathbf{y}_{m,\mathrm{d}}^{(t)} &= \sum_{k\in\mathcal{A}} \mathbf{g}_{k,m,\mathrm{d}} x_{k,\mathrm{d}}^{(t)} + \mathbf{z}_{m,\mathrm{d}}^{(t)} \\ &= \mathbf{G}_{m,\mathrm{d}}\mathbf{x}_{\mathrm{d}}^{(t)} + \mathbf{z}_{m,\mathrm{d}}^{(t)} \in \mathbb{C}^{D\times 1},\end{aligned} \tag{5.6}$$

where $\mathbf{g}_{k,m,\mathrm{d}} \in \mathbb{C}^{D\times 1}$ and $\mathbf{z}_{m,\mathrm{d}}^{(t)} \sim \mathcal{CN}(\mathbf{0}_D, \sigma_{\mathrm{n}}^2\mathbf{I}_D)$ denote the CFRs between the BS and user k at the m-th receiving antenna and the AWGN, respectively. In addition, $x_{k,\mathrm{d}}^{(t)} \in \mathcal{X}$, where $\mathcal{X}$ is the set of Q-ary modulated symbols, represents the t-th data symbol transmitted by user k. The matrix $\mathbf{G}_{m,\mathrm{d}} \in \mathbb{C}^{D\times K_{\mathrm{a}}}$ and the vector $\mathbf{x}_{\mathrm{d}}^{(t)} \in \mathcal{X}^{K_{\mathrm{a}}\times 1}$ comprise the active users' CFRs and data symbols, respectively. For ease of data transmission, the transmitted data symbols are directly mapped onto subcarriers in subset $\mathcal{D}$, which comprises the indices that are uniformly picked from all available subcarriers with the exception of $\mathcal{P}$.

The signals received by M antennas can be expressed as

$$\begin{aligned}\mathbf{y}_{\mathrm{d}}^{(t)} &= [(\mathbf{y}_{1,\mathrm{d}}^{(t)})^{\mathrm{T}}, \ldots, (\mathbf{y}_{M,\mathrm{d}}^{(t)})^{\mathrm{T}}]^{\mathrm{T}} \\ &= \begin{bmatrix}\mathbf{G}_{1,\mathrm{d}} \\ \vdots \\ \mathbf{G}_{M,\mathrm{d}}\end{bmatrix} \mathbf{x}_{\mathrm{d}}^{(t)} + \begin{bmatrix}\mathbf{z}_{1,\mathrm{d}}^{(t)} \\ \vdots \\ \mathbf{z}_{M,\mathrm{d}}^{(t)}\end{bmatrix} = \mathbf{G}_{\mathrm{d}}\mathbf{x}_{\mathrm{d}}^{(t)} + \mathbf{Z}_{\mathrm{d}}^{(t)} \in \mathbb{C}^{DM\times 1}.\end{aligned} \tag{5.7}$$

Note that, as is true for the pilot component, the CFRs in (5.7) can be obtained using the CIRs , *i.e.,*

$$[\mathbf{g}_{k,1,\mathrm{d}}, \ldots, \mathbf{g}_{k,M,\mathrm{d}}] = \sqrt{N}\mathbf{F}_{D,L}\mathbf{H}_k,\ k \in \mathcal{A}, \tag{5.8}$$

where $\mathbf{F}_{D,L} \in \mathbb{C}^{D\times L}$ is a matrix comprising D rows according to $\mathcal{D}$ and the first L columns of $\mathbf{F}_N$. This enables efficient MUD based on the results of AUD and CE.

5.3 Sequence and System Designs for Low-Latency Massive Grant-Free Access

In this section, we introduce a sequence and system design for satisfying both low latency and massive connectivity requirements. To achieve this, the proposed method replaces the DCS theory-based sequence design in [64] with an approach in which different unimodular sequences are applied to different OFDM symbols to enlarge the dimensionality of $\mathbf{Y}_{\mathrm{p}}$. We then propose a suitable design through which the proposed GF-NOMA system can meet heterogeneous requirements while conforming to 5G generation new radio (5G NR).

5.3.1 Sequence Design for Massive Access

In the CS problem given by (5.5), the matrix $\mathbf{A}$ serves as a measurement matrix whose properties are crucial for the accuracy of sparse recovery and, therefore, for which the sequences should be carefully designed.

In related studies *e.g.,* [54, 69, 63], it was assumed that all sequences are generated from an i.i.d. complex Gaussian distribution with zero mean. The JACE scheme presented in this chapter is based on the GMMV-AMP algorithm [63, 64], which is a member of the family of AMP algorithms that are theoretically guaranteed to reach fixed points when i.i.d. Gaussian measurement matrices are employed [78, 104]. However, the measurement matrices applied in those studies assumed $\mathbf{A}$ to be composed solely of predesigned sequences, whereas as shown in (5.5) $\mathbf{A}$ is in fact the product of $\mathrm{diag}(\mathbf{s}_k^{(t)})$ and $\bar{\mathbf{F}}_{P,L}$.

Hence, as noted in Section 5.2, we employ a random unimodular sequence, *i.e.,* $|s_{k,p}^{(t)}| = 1$, like [62, 94]. We note that the gram matrix of $\mathbf{A}_k^{(t)}$, namely $(\mathbf{A}_k^{(t)})^{\mathrm{H}}\mathbf{A}_k^{(t)} \in \mathbb{C}^{L\times L}$, is equal to the identity matrix $\mathbf{I}_L$ if N/P is an integer. This leads to the orthogonality of any pairs of different columns in $\mathbf{A}_k^{(t)}$. In our approach, different sequences are used for different OFDM symbols to enlarge the dimensionality of $\mathbf{Y}_{\mathrm{p}}$ and improve the estimation accuracy.

It should be noted that the off-diagonal elements of the gram matrix of $\mathbf{A}$, *i.e.*, $\mathbf{G}_{\mathbf{A}} = \mathbf{A}^{\mathrm{H}}\mathbf{A}$, can be regarded as sums of random phase vectors and therefore, as a result of the central limit theorem, tend to be complex Gaussian random variables with zero means and variances of $1/PT$ for large values of $\mathbf{A}$ [105]. Indeed, this is seen from Fig. 5.3, which shows the PDF and corresponding Rayleigh distribution of the

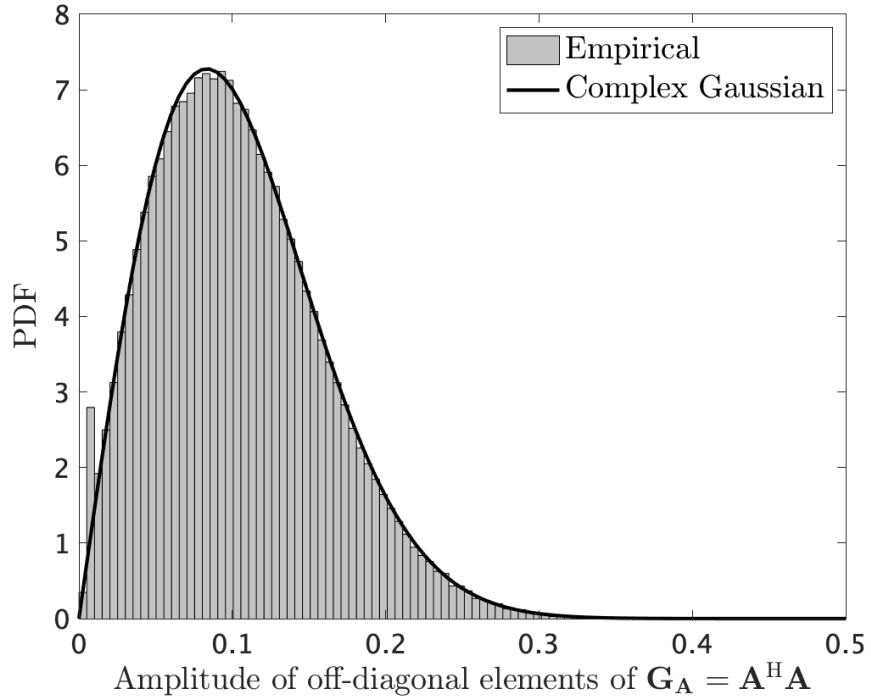

Fig. 5.3 Example of PDF of amplitudes of off-diagonal elements in the gram matrix $\mathbf{G}_\mathbf{A} = \mathbf{A}^\mathrm{H}\mathbf{A}$ where $K = 100$, $L = 25$, $P = 36$, and $T = 2$.

amplitudes of the off-diagonal elements of $\mathbf{G}$. In addition, the matrix $\mathbf{G}$ tends to have many near-zero entries because $(\mathbf{A}_k^{(t)})^\mathrm{H}\mathbf{A}_k^{(t)}$ will be similar to the identity matrix.

5.3.2 Phase Transition-Based System Design

In CS-based approaches, the relation between the activity and undersampling ratios influences the sparse recovery performance. The *phase transition* occurring in the family of AMP algorithms has been investigated to clarify the sparsity-measurement trade-off of these algorithms in terms of the variables (ρ, δ) [85, 106–108], which are given by $\rho = k/m$ and $\delta = m/n$ for $\rho \in [0, 1]$ and $\delta \in [0, 1]$, respectively, where k is the number of non-zero elements (rows) in the desired sparse signal and the measurement matrix is an $m \times n$ matrix. In practice, GF-NOMA systems should be carefully designed to meet low latency and massive connectivity requirements by taking into account system parameters such as subcarrier spacing and system bandwidth.

Therefore, we propose a novel system design based on analyses of both ℓ_1-norm minimization in the complex domain [107] and the AMP algorithm for block-sparse recovery [106]. Although empirical phase transitions in recovering Bernoulli-Gaussian

signals have been demonstrated for both the generalized AMP algorithm using the EM [85] and the orthogonal AMP [108], here we apply theoretical phase transitions to ensure a simple and flexible design.

Theoretical phase transition can be obtained by evaluating the minimax mean-squared error (MSE), which is defined as $\mathcal{M}(\epsilon|\eta)$, where $\epsilon = \rho\delta$ and η denote, respectively, a sparsity ratio and a denoiser for AMP [106]. Then, AMP succeeds with high probability when the following condition is satisfied [106]:

$$\delta > \mathcal{M}(\epsilon|\eta). \tag{5.9}$$

As a successful region, we consider the phase transition derived in [107], in which an SMVR problem in the complex domain is considered. Because this is a special case of an MMVR problem, its estimation performance can serve as the lower bound of that of an MMVR problem. According to [107, Theorem III.5], ρ and δ for AMP using the complex-valued soft-thresholding denoiser, *i.e.,* complex AMP (CAMP), satisfy the following relation for $\tau \in [0, \infty)$:

$$\rho = \frac{\chi_1(\tau)}{(1+\tau^2)\chi_1(\tau) - \tau\chi_2(\tau)} \tag{5.10}$$

$$\delta = \frac{4(1+\tau^2)\chi_1(\tau) - 4\tau\chi_2(\tau)}{-2\tau + 4\chi_2(\tau)}, \tag{5.11}$$

where

$$\chi_1(\tau) \triangleq \int_{\omega \geq \tau} \omega(\tau - \omega)\mathrm{e}^{-\omega^2} d\omega, \tag{5.12}$$

$$\chi_2(\tau) \triangleq \int_{\omega \geq \tau} \omega(\tau - \omega)^2\mathrm{e}^{-\omega^2} d\omega. \tag{5.13}$$

The largest phase transition for CAMP can then be obtained by exploiting the τ that maximizes the value of ρ in (5.10).

In contrast, we exploit a theoretical approach to recovering block-sparse signals [106] to obtain the largest achievable phase transition by taking advantage of the fact that MMVR problems can be expressed using SMVR problems with block-sparse signals. This approach takes the performance of AMP into consideration using the block soft-thresholding denoiser. To clarify the ultimate boundary, we focus on the case of an infinite block size corresponding to $M \to \infty$. Following [106, Lemma 3.3],

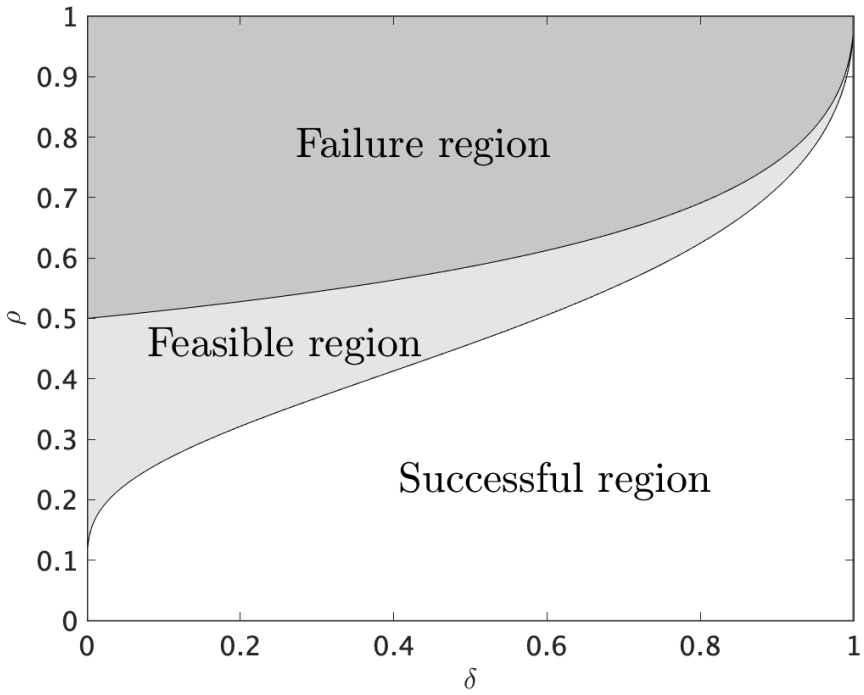

Fig. 5.4 Illustration of regions classified by phase transitions in [106] and [107], providing guidelines for proper parameter design.

we obtain the minimax MSE for a large block size:

$$\mathcal{M}(\epsilon|\eta) = 2\epsilon - \epsilon^2. \tag{5.14}$$

According to (5.9) and (5.14), the largest phase transition for a large block size is given by $\delta = 2\epsilon - \epsilon^2$.

From these theoretically derived phase transitions, we can obtain a guideline for proper parameter design, as indicated by the three regions in Fig. 5.4, which are divided by boundaries obtained from the largest phase transition for CAMP and using (5.14) with $M \to \infty$. In Fig. 5.4, the "Failure region" represents the region in which the accurate estimation of $\mathbf{X}$ using (5.5) is absolutely impossible even if $M \to \infty$, whereas the "Successful region" is where the highest degree of accuracy of estimation is surely achievable and the "Feasible region" is where accurate estimation can be performed because it has a boundary for $M > 1$. Thus, the figure indicates that the values of ρ and δ in the proposed GF-NOMA system should exist in at least the "Feasible region."

Throughout the remainder of this chapter, the variables ρ and δ in the proposed GF-NOMA are given for the sake of simplicity by

$$\rho = \frac{K_{\mathrm{a}} L_{\mathrm{path}}}{PT}, \tag{5.15}$$

$$\delta = \frac{PT}{KL}, \tag{5.16}$$

where $K_{\mathrm{a}} L_{\mathrm{path}}$ corresponds to the maximum value of the number of non-zero rows in $\mathbf{X}$, which varies as a result of the randomness of path delays. The values of T that satisfy the latency requirement are strictly limited by the subcarrier spacing, ΔB, and the value of L depends on both ΔB and the system bandwidth B_{s}. Accordingly, we can use the number of pilot subcarriers and OFDM symbols for AUD and CE to determine the successful (or feasible) region characterized by phase transitions taking into account ΔB and B_{s}.

5.4 AUD, CE, and MUD for Massive GF-NOMA

In this section, we describe the AUD, CE, and MUD schemes used by the proposed GF-NOMA. After describing the joint AUD and CE scheme, which is based on the GMMV-AMP algorithm [64], we present the MUD scheme, which is implemented via the symbol-wise GaBP [101, 102].

5.4.1 Preliminaries

The GMMV-AMP algorithm solves the CS problem in (5.5) and approximately calculates an MMSE estimate of $\mathbf{X}$ as the posterior mean, which can be expressed as

$$\hat{x}_{k,m} = \int x_{k,m} p(x_{k,m}|\mathbf{Y}_{\mathrm{p}}) dx_{k,m}, \ \forall k, m, \tag{5.17}$$

where $p(x_{k,m}|\mathbf{Y}_{\mathrm{p}})$ is the marginal posterior distribution given by

$$p(x_{k,m}|\mathbf{Y}_{\mathrm{p}}) = \int p(\mathbf{X}|\mathbf{Y}_{\mathrm{p}}) d\mathbf{X}_{\setminus k,m}, \tag{5.18}$$

where $\mathbf{X}_{\backslash k,m}$ denotes the collection of elements of $\mathbf{X}$ aside from $x_{k,m}$. Based on the Bayesian rule, the joint posterior distribution in (5.18) can be rewritten as

$$\begin{aligned} p(\mathbf{X}|\mathbf{Y}_{\mathrm{p}}) &= \frac{p(\mathbf{Y}_{\mathrm{p}}|\mathbf{X})p_0(\mathbf{X})}{p(\mathbf{Y}_{\mathrm{p}})} \\ &= \frac{1}{\tilde{Z}_1}\prod_{m=1}^{M}\left[\prod_{\ell=1}^{PT} p(y_{\ell,m}|\mathbf{X})\prod_{k=1}^{KL} p_0(x_{k,m})\right], \end{aligned} \tag{5.19}$$

where $\tilde{Z}_1$ is a normalization factor and $p_0(\mathbf{X})$ is the a priori distribution of $\mathbf{X}$. Under the AWGN assumption[1], the likelihood function $p(y_{\ell,m}|\mathbf{X})$ can be computed as

$$p(y_{\ell,m}|\mathbf{X}) = \frac{1}{\pi\sigma_{\mathrm{n}}^2}\exp\left(-\frac{1}{\sigma_{\mathrm{n}}^2}\left|y_{\ell,m} - \sum_{k} a_{\ell,k}x_{k,m}\right|^2\right), \tag{5.20}$$

where $a_{\ell,k}$ is the (ℓ,k)-th entry of $\mathbf{A}$ in (5.5). The a priori distribution of $\mathbf{X}$ is assumed to be a Bernoulli-Gaussian distribution, *i.e.,*

$$\begin{aligned} p_0(\mathbf{X}) &= \prod_{m=1}^{M}\prod_{k=1}^{KL} p_0(x_{k,m}) \\ &= \prod_{m=1}^{M}\prod_{k=1}^{KL}[(1-\gamma_{k,m})\delta(x_{k,m}) + \gamma_{k,m}\mathcal{CN}(0,\nu_{k,m})], \end{aligned} \tag{5.21}$$

where $\gamma_{k,m}$ is the sparsity ratio and $\delta(\cdot)$ is the Dirac delta function. Note that $\gamma_{k,m}$ and $\nu_{k,m}$ are hyperparameters, which are learned using the EM algorithm to avoid the difficulty of obtaining the full knowledge of the a priori distribution of the channels in advance. Unlike the system in [64], the proposed GF-NOMA system attempts to obtain the CIRs accurately, which allows us to assume that the mean of the channel gains is zero.

5.4.2 Message Passing Procedures

The proposed scheme iterates the updating of messages at the variable and function nodes of the bipartite graph representing the factorization in (5.19). To avoid difficulties in computing (5.18), $p(x_{k,m}|\mathbf{Y}_{\mathrm{p}})$ is approximated based on a decoupling of the estimation problem of $\mathbf{X}$ into KLM scalar estimation problems. Thus, the posterior distribution

[1]Although the noise variance is considered to be an unknown parameter in [64], its value is assumed to be usable at the BS in this chapter.

of $x_{k,m}$ can be approximated as

$$\begin{aligned} p(x_{k,m}|\mathbf{Y}_{\mathrm{p}}) &\approx p(x_{k,m}|C_{k,m}^{(q)}, D_{k,m}^{(q)}) \\ &\approx \frac{1}{\tilde{Z}_2} p_0(x_{k,m}) \mathcal{CN}(C_{k,m}^{(q)}, D_{k,m}^{(q)}), \end{aligned} \tag{5.22}$$

where q and $\tilde{Z}_2$ are the iteration index of the algorithm and a normalization factor, respectively, and $C_{k,m}^{(q)}$ and $D_{k,m}^{(q)}$ are calculated at the variable nodes of the bipartite graph.

Following Proposition 1 in [64], the following updates are first executed at the function nodes of the bipartite graph to obtain the values of $C_{k,m}^{(q)}$ and $D_{k,m}^{(q)}$ in (5.22):

$$V_{\ell,m}^{(q)} = \sum_{k} |a_{\ell,k}|^2 v_{k,m}^{(q)}, \tag{5.23}$$

$$Z_{\ell,m}^{(q)} = \sum_{k} a_{\ell,k} \hat{x}_{k,m}^{(q)} - \frac{V_{\ell,m}^{(q)}}{\sigma_{\mathrm{n}}^2 + V_{\ell,m}^{(q-1)}} (y_{\ell,m} - Z_{\ell,m}^{(q-1)}), \tag{5.24}$$

where $\hat{x}_{k,m}^{(q)}$ and $v_{k,m}^{(q)}$ are the posterior mean and variance at the q-th iteration, respectively. Following this, $C_{k,m}^{(q)}$ and $D_{k,m}^{(q)}$ in (5.22) are updated at the variable nodes of the bipartite graph as follows:

$$D_{k,m}^{(q)} = \left[\sum_{\ell} \frac{|a_{\ell,k}|^2}{\sigma_{\mathrm{n}}^2 + V_{\ell,m}^{(q)}} \right]^{-1}, \tag{5.25}$$

$$C_{k,m}^{(q)} = \hat{x}_{k,m}^{(q)} + D_{k,m}^{(q)} \sum_{\ell} \frac{a_{\ell,k}^* (y_{\ell,m} - Z_{\ell,m}^{(q)})}{\sigma_{\mathrm{n}}^2 + V_{\ell,m}^{(q)}}. \tag{5.26}$$

For more details on the derivation of (5.23)–(5.26), please refer to Appendix A in [64].

We next consider the updating of the posterior mean and variance. Under the assumption of an a priori distribution of channel gains, *i.e.,* $x_{k,m} \sim \mathcal{CN}(0, \nu_{k,m})$, and applying the model in (5.22), the posterior distribution of $x_{k,m}$ can be obtained as follows[2]:

$$p(x_{k,m}|C_{k,m}^{(q)}, D_{k,m}^{(q)}) = (1 - \xi_{k,m}^{(q)}) \delta(x_{k,m}) + \xi_{k,m}^{(q)} \mathcal{CN}(A_{k,m}^{(q)}, B_{k,m}^{(q)}), \tag{5.27}$$

[2]This can be derived via the Gaussian-PDF multiplication rule. Please refer to [85] for further details.

where

$$A_{k,m}^{(q)} = \frac{\nu_{k,m} C_{k,m}^{(q)}}{D_{k,m}^{(q)} + \nu_{k,m}}, \tag{5.28}$$

$$B_{k,m}^{(q)} = \frac{\nu_{k,m} D_{k,m}^{(q)}}{\nu_{k,m} + D_{k,m}^{(q)}}, \tag{5.29}$$

$$\xi_{k,m}^{(q)} = \frac{\gamma_{k,m}}{\gamma_{k,m} + (1 - \gamma_{k,m}) \exp(-\Lambda)}, \tag{5.30}$$

$$\Lambda = \ln \frac{D_{k,m}^{(q)}}{D_{k,m}^{(q)} + \nu_{k,m}} + \frac{|C_{k,m}^{(q)}|^2}{D_{k,m}^{(q)}} - \frac{|C_{k,m}^{(q)}|^2}{D_{k,m}^{(q)} + \nu_{k,m}}, \tag{5.31}$$

and $\xi_{k,m}^{(q)}$ represents the belief indicator at the q-th iteration. The posterior mean and variance can then be explicitly updated as

$$\hat{x}_{k,m}^{(q+1)} = \xi_{k,m}^{(q)} A_{k,m}^{(q)}, \tag{5.32}$$

$$v_{k,m}^{(q+1)} = \xi_{k,m}^{(q)} \left(\left| A_{k,m}^{(q)} \right|^2 + B_{k,m}^{(q)} \right) - \left| \hat{x}_{k,m}^{(q+1)} \right|^2, \tag{5.33}$$

respectively. Based on these updates, the MMSE estimate of $\mathbf{X}$ can be acquired through iteration of (5.23)–(5.33).

5.4.3 Update of Hyperparameters via EM Algorithm

The EM is used, in conjunction with the procedures described in the previous subsection, to learn the hyperparameters, *i.e.*, $\boldsymbol{\theta} = \{\nu_{k,m}, \gamma_{k,m}, \forall k, m\}$. This enables the accurate estimation without knowing the a priori channel distribution. Note that, at the q-th iteration, the variables $\nu_{k,m}$ and $\gamma_{k,m}$ are replaced by $\nu_{k,m}^{(q)}$ and $\gamma_{k,m}^{(q)}$, respectively. Moreover, we adopt the following hyperparameter update rules:

$$\gamma_{k,m}^{(q+1)} = \xi_{k,m}^{(q+1)} = \frac{1}{M} \sum_{m=1}^{M} \frac{\gamma_{k,m}^{(q)}}{\gamma_{k,m}^{(q)} + (1 - \gamma_{k,m}^{(q)}) \exp(-\Lambda)}, \tag{5.34}$$

$$\nu_{k,m}^{(q+1)} = \frac{\sum_k \xi_{k,m}^{(q)} \left(|A_{k,m}^{(q)}|^2 + B_{k,m}^{(q)} \right)}{\sum_k \xi_{k,m}^{(q)}}. \tag{5.35}$$

Because $\mathbf{X}$ has row-sparsity, the update rule for $\gamma_{k,m}^{(q+1)}$ is based on [63]. In contrast, the update rule for $\nu_{k,m}^{(q+1)}$ differs slightly from that used in [64] because it is assumed that the mean of the channel gains, *i.e.*, μ in [64], is zero.

Algorithm 5.1 GMMV-AMP-based AUD and CE

Input: Received signals $\mathbf{Y}_{\mathrm{p}} \in \mathbb{C}^{PT \times M}$, measurement matrix $\mathbf{A} \in \mathbb{C}^{PT \times KL}$, damping factor ζ, maximum number of iterations T_{amp}, and termination threshold η_{th}.

1: Initialize the iteration index q to 1, the hyperparameters as in (5.36) and (5.37), and the other parameters as $V_{\ell,m}^{(0)} = 1$, $Z_{\ell,m}^{(0)} = y_{\ell,m}$, $\hat{x}_{k,m}^{(1)} = 0$, and $v_{k,m}^{(1)} = \tau_{k,m}^{(1)}$, $\forall \ell, k, m$, respectively.

2: **repeat**

3: At the function nodes, obtain the messages $V_{\ell,m}^{(q)}$ and $Z_{\ell,m}^{(q)}$ using (5.23) and (5.24), respectively, $\forall \ell, m$.

4: Update $V_{\ell,m}^{(q)}$ and $Z_{\ell,m}^{(q)}$ with damping, $\forall \ell, m$:
$V_{\ell,m}^{(q)} = \zeta V_{\ell,m}^{(q-1)} + (1-\zeta) V_{\ell,m}^{(q)}$,
$Z_{\ell,m}^{(q)} = \zeta Z_{\ell,m}^{(q-1)} + (1-\zeta) Z_{\ell,m}^{(q)}$.

5: At the variable nodes, update the messages $D_{k,m}^{(q)}$ and $C_{k,m}^{(q)}$ using (5.25) and (5.26), respectively, $\forall k, m$.

6: Obtain the posterior mean $\hat{x}_{k,m}^{(q+1)}$ and variance $v_{k,m}^{(q+1)}$ using (5.32) and (5.33), respectively, $\forall k, m$.

7: Update the hyperparameters $\gamma_{k,m}^{(q+1)}$ and $\nu_{k,m}^{(q+1)}$ using (5.34) and (5.35), respectively, $\forall k, m$.

8: $q = q + 1$.

9: **until** $q > T_{\mathrm{amp}}$ or $\|\hat{\mathbf{X}}^{(q)} - \hat{\mathbf{X}}^{(q-1)}\|_{\mathrm{F}} / \|\hat{\mathbf{X}}^{(q-1)}\|_{\mathrm{F}} < \eta_{\mathrm{th}}$.

Output: The estimate of $\mathbf{X}$; $\hat{\mathbf{X}}^{(q)}$ and the belief indicators $\gamma_{k,m}^{(q)}$, $\forall k, m$.

Since the initialization of the unknown parameters has an impact on the performance of the above EM update, we use the following initialization for the hyperparameters:

$$\tau_{k,m}^{(1)} = \frac{\sum_{\ell} |y_{\ell,m}|^2 - M\sigma_{\mathrm{n}}^2}{\sum_{\ell} \sum_{k} |a_{\ell,k}|^2}, \tag{5.36}$$

$$\gamma_{k,m}^{(1)} = \frac{1}{2}. \tag{5.37}$$

Notice that we use (5.37) owing to the randomness of the sparsity level of $\mathbf{X}$ in (5.5).

5.4.4 Active User Detection and Channel Estimation

The resulting overall algorithm is summarized as Algorithm 5.1. Note that, in line 4 of Algorithm 5.1, we apply a damping with a damping factor of $\zeta \in [0, 1]$ to avoid divergence of the algorithm. It is further worth noting that the computational complexity for each iteration in this algorithm is $\mathcal{O}(KLMPT)$, which increases linearly with K, L, M, P, and T because matrix inversion is avoided.

After processing Algorithm 5.1, an estimate of the channels is obtained from the estimate of $\mathbf{X}$, namely $\hat{\mathbf{X}}^{(q)}$. Moreover, the estimated set of active users is determined based on $\gamma_{k,m}^{(q)}$:

$$\begin{cases} k \in \hat{\mathcal{A}}, & \text{if } \gamma_{k,\max} > \frac{1}{2} \\ k \notin \hat{\mathcal{A}}, & \text{otherwise,} \end{cases} \tag{5.38}$$

where

$$\gamma_{k,\max} \triangleq \max_{\ell=1,\ldots,L} \gamma_{k(L-1)+\ell}^{(q)}. \tag{5.39}$$

5.4.5 Multiuser Detection via GaBP

As the estimated CFRs in the data component must be also able to estimate the transmitted data symbols, they are recovered based on the relation between the CFRs and CIRs, *i.e.,*

$$[\hat{\mathbf{g}}_{k,1,\mathrm{d}}, \ldots, \hat{\mathbf{g}}_{k,m,\mathrm{d}}] = \sqrt{N}\mathbf{F}_{D,L}\hat{\mathbf{H}}_k, \ k \in \hat{\mathcal{A}}. \tag{5.40}$$

Eventually, an estimate of $\mathbf{G}_{\mathrm{d}}$ in (5.7), denoted by $\hat{\mathbf{G}}_{\mathrm{d}}$, of size $DM \times \hat{K}_{\mathrm{a}}$, where $\hat{K}_{\mathrm{a}}$ denotes the cardinality of $\hat{\mathcal{A}}$, is obtained.

The proposed GF-NOMA estimates transmitted data symbols by applying the symbol-wise GaBP algorithm utilizing $\mathbf{y}_{\mathrm{d}}^{(t)}$ and $\hat{\mathbf{G}}_{\mathrm{d}}$. This algorithm iterates the following operations: 1) soft interference cancellation, 2) updating the prior belief based on a scalar Gaussian approximation, 3) updating the extrinsic belief, and 4) calculation of the soft symbol based on the extrinsic belief. Hereafter, the superscript t and the subscript d in $\mathbf{y}_{\mathrm{d}}^{(t)}$, $\hat{\mathbf{G}}_{\mathrm{d}}$, $\mathbf{x}_{\mathrm{d}}^{(t)}$, and $\mathbf{z}_{\mathrm{d}}^{(t)}$ are dropped to simplify the notation.

In the q-th iteration, the soft interference cancellation is first carried out using the soft symbol vector obtained from the previous iteration, *i.e.,* $\tilde{\mathbf{x}}_{\ell}^{(q-1)}$. The ℓ-th received symbol y_ℓ following soft interference cancellation can then be expressed as

$$\tilde{y}_{\ell,k}^{(q)} = y_\ell - \hat{\mathbf{g}}_\ell \tilde{\mathbf{x}}_{\ell}^{(q-1)} + \hat{g}_{\ell,k}\tilde{x}_{\ell,k}^{(q-1)}, \tag{5.41}$$

where $\hat{\mathbf{g}}_\ell \in \mathbb{C}^{1\times \hat{K}_{\mathrm{a}}}$ denotes the ℓ-th row of $\hat{\mathbf{G}}_{\mathrm{d}}$.

To relax the calculation of the beliefs, (5.41) is modeled as a simple Gaussian model using scalar Gaussian approximation:

$$\tilde{y}_{\ell,k}^{(q)} = \hat{g}_{\ell,k}x_k + \nu_{\ell,k}^{(q)}, \tag{5.42}$$

with the effective noise $\nu_{\ell,k}^{(q)} \sim \mathcal{CN}(0, \psi_{\ell,k}^{(q)})$. The conditional variance $\psi_{\ell,k}^{(q)}$ is given by

$$\psi_{\ell,k}^{(q)} = \hat{\mathbf{g}}_\ell \boldsymbol{\Delta}_{\ell,k}^{(q)} \hat{\mathbf{g}}_\ell^{\mathrm{H}} + \sigma_{\mathrm{n}}^2, \tag{5.43}$$

where

$$\boldsymbol{\Delta}_{\ell,k}^{(q)} = \mathrm{diag}\left(\delta_{\ell,1}^{(q)}, \ldots, \delta_{\ell,k-1}^{(q)}, 0, \delta_{\ell,k+1}^{(q)}, \ldots, \delta_{\ell,\hat{K}_{\mathrm{a}}}^{(q)}\right), \tag{5.44}$$

$$\delta_{\ell,k}^{(q)} = \mathbb{E}_{x_k|\boldsymbol{\beta}_{\ell,k}^{(q-1)}}[|x_k|^2] - |\tilde{x}_{\ell,k}^{(q-1)}|^2. \tag{5.45}$$

The first term in (5.45) is calculated based on the extrinsic belief $\beta_{\ell,k}^{(q-1)}(x_s)$:

$$\mathbb{E}_{x_k|\boldsymbol{\beta}_{\ell,k}^{(q-1)}}[|x_k|^2] = \sum_{x_s \in \mathcal{X}} |x_s|^2 \frac{\exp\left(\beta_{\ell,k}^{(q-1)}(x_s)\right)}{\sum\limits_{\tilde{x}_s \in \mathcal{X}} \exp\left(\beta_{\ell,k}^{(q-1)}(\tilde{x}_s)\right)}. \tag{5.46}$$

Under the scalar Gaussian approximation, the PDF of $\tilde{y}_{\ell,k}$ for $x_s \in \mathcal{X}$ can be expressed as

$$\begin{aligned} p(\tilde{y}_{\ell,k}^{(q)}|x_s &\propto \exp\left(-\frac{|\tilde{y}_{\ell,k}^{(q)} - \hat{g}_{\ell,k} x_s|^2}{\psi_{\ell,k}^{(q)}}\right) \\ &\propto \exp\left(\omega_{\ell,k}^{(q)}(x_s)\right), \end{aligned} \tag{5.47}$$

where

$$\omega_{\ell,k}^{(q)}(x_s) = \frac{2\Re\{x_s^* \hat{g}_{\ell,k}^* \tilde{y}_{\ell,k}^{(q)}\} - |\hat{g}_{\ell,k}|^2 |x_s|^2}{\psi_{\ell,k}^{(q)}}. \tag{5.48}$$

Eventually, the prior belief $\vartheta_{\ell,k}^{(q)}(x_s)$ is updated under the above belief as follows:

$$\vartheta_{\ell,k}^{(q)}(x_s) = \tilde{\zeta} \vartheta_{\ell,k}^{(q-1)}(x_s) + (1 - \tilde{\zeta}) \omega_{\ell,k}^{(q)}(x_s), \tag{5.49}$$

in which the damping factor $\tilde{\zeta} \in [0, 1]$ suppresses the negative effects of the approximation error [109, 110].

The extrinsic belief $\beta_{\ell,k}^{(q)}(x_s)$ is then updated as follows:

$$\beta_{\ell,k}^{(q)}(x_s) = \sum_{\ell'=1, \ell' \neq \ell}^{DM} \vartheta_{\ell',k}^{(q)}(x_s), \tag{5.50}$$

Algorithm 5.2 GaBP-based MUD

Input: Received signal $\mathbf{y}_{\mathrm{d}}^{(t)} \in \mathbb{C}^{MD\times 1}$, estimated CFRs $\hat{\mathbf{G}}_{\mathrm{d}} \in \mathbb{C}^{MD\times \hat{K}_{\mathrm{a}}}$, damping factor $\tilde{\zeta}$, and number of iterations T_{gabp}.

Initialize the iteration index q to 1 and the other variables as $\tilde{x}_{\ell,k}^{(0)} = 0$, $\vartheta_{k,m}^{(0)}(x_s) = 0$, and $\beta_{k,m}^{(0)}(x_s) = 0$, $\forall \ell, k, x_s$, respectively.
for $q = 1, 2, \ldots, T_{\mathrm{gabp}}$ **do**
  for $\ell = 1, 2, \ldots, DM$ **do**
    for $k = 1, 2, \ldots, \hat{K}_{\mathrm{a}}$ **do**
      Perform soft interference cancellation using (5.41).
      Update the variance of $\nu_{\ell,k}^{(q)}$ using (5.43).
      Update the prior belief $\vartheta_{\ell,k}^{(q)}(x_s)$ using (5.49).
      Update the extrinsic belief $\beta_{\ell,k}^{(q)}(x_s)$ using (5.50).
      Calculate the soft symbol $\tilde{x}_{\ell,k}^{(q)}$ using (5.51).
    end for
  end for
end for
for $k = 1, 2, \ldots, \hat{K}_{\mathrm{a}}$ **do**
  Determine the estimated data symbol $\hat{x}_{k,\mathrm{d}}^{(t)}$ using (5.52).
end for

Output: The estimate of data symbols $\hat{x}_{k,\mathrm{d}}^{(t)}$.

and then the soft symbol $\tilde{x}_{\ell,k}^{(q)}$ is calculated based on the extrinsic belief for the next iteration:

$$\tilde{x}_{\ell,k}^{(q)} = \sum_{x_s\in\mathcal{X}} x_s \frac{\exp\left(\beta_{\ell,k}^{(q)}(x_s)\right)}{\sum\limits_{\tilde{x}_s\in\mathcal{X}} \exp\left(\beta_{\ell,k}^{(q)}(\tilde{x}_s)\right)}. \tag{5.51}$$

After T_{gabp} iterations of (5.41)–(5.51), the estimated data symbols are determined as follows:

$$\hat{x}_{k,\mathrm{d}}^{(t)} = \arg\max_{x_s\in\mathcal{X}} \sum_{\ell=1}^{DM} \vartheta_{\ell,k}^{(q)}(x_s). \tag{5.52}$$

The overall MUD process applied by the proposed GF-NOMA is summarized as Algorithm 5.2.

In the MUD process, the computational complexity of estimating the CFRs in the data component is $\mathcal{O}(D\hat{K}_{\mathrm{a}}LM)$ (based on (5.40)), whereas the complexity of data detection by the symbol-wise GaBP is $\mathcal{O}(D\hat{K}_{\mathrm{a}}MQT_{\mathrm{gabp}})$. As the CFRs are estimated

Table 5.1 Computational complexity of JACE. Here, the superscripts † and ‡ indicate the case in [64] and $T = 1$, respectively.

Scheme	Complexity order in each iteration
GMMV-AMP with model (5.5)	$\mathcal{O}(KLMPT)$
GMMV-AMP [64]†	$\mathcal{O}(KMPT)$
Turbo-GMMV-AMP [64]†	$\mathcal{O}(T_{\mathrm{amp}}KMPT)$
MRAS [62]‡	$\mathcal{O}(KLMP)$ (if $M < P$)

only once over the duration of a single subframe, the computational complexity of the MUD process is primarily driven by the processing of the symbol-wise GaBP.

5.4.6 Overall Computational Complexity

The proposed GF-NOMA performs JACE using GMMV-AMP and MUD using GaBP continuously. As discussed in Sections 5.4.4 and 5.4.5, the overall computational complexity over the duration of a single subframe is given by $\mathcal{O}(MT(KLPT_{\mathrm{amp}} + D\hat{K}_{\mathrm{a}}N_{\mathrm{sym}}QT_{\mathrm{gabp}}))$, where N_{sym} is the number of transmissible data symbols in a single OFDM symbol.

For comparison, TABLE 5.1 lists the computational complexity of the JACE scheme under the model (5.5) along with those of state-of-the-art JACE schemes for GF-NOMA with MIMO-OFDM [64, 62] in terms of the complexity order per iteration. We note that the complexity order of the proposed scheme is generally higher than those of the state-of-the-art schemes owing to the differences between the signal models used. In fact, applying the scheme proposed in [62] to the model (5.5), *i.e.,* $P \to PT$ $(T > 1)$, reveals that, in terms of computational complexity, the proposed scheme is comparable to MRAS [62]. The complexity order of the proposed scheme is also comparable to that of turbo-GMMV-AMP. These results indicate that the proposed scheme is computationally efficient for massive access scenarios.

5.5 Numerical Results

We evaluated the performance of the proposed GF-NOMA system using computer simulation. The simulation parameter values are listed in Table 5.2. Note that $T = 28$ is equivalent to the number of transmissible OFDM symbols of length 1 msec within a 5G NR subframe with 30 kHz subcarrier spacing. As each resource block (RB) in 5G NR comprises 12 subcarriers, we based the number of pilot subcarriers P on the assumption that RB is one unit. In all simulations, the SNR was defined as $\mathrm{SNR} \triangleq 1/\sigma_{\mathrm{n}}^2$

Table 5.2 Simulation parameters

Number of potential users K	500
Number of active users K_{a}	50,100
Number of antennas at the BS M	8
Number of subcarriers N	4096
Number of pilot subcarriers P	$12 \times$ Number of RBs
Number of subcarriers for data D	16,18
Number of used OFDM symbols T	28
Number of (visible) taps in CIRs L	25
Number of significant path L_{path}	6
Subcarrier spacing ΔB	30 kHz
System bandwidth B_{s}	10 MHz
Modulation order Q	4
Damping factor of GMMV-AMP ζ	0.3
Number of maximum iterations of GMMV-AMP T_{amp}	200
Termination threshold of GMMV-AMP η_{th}	10^{-5}
Damping factor of GaBP $\tilde{\zeta}$	0.5
Number of iterations of GaBP T_{gabp}	16

and Gray-coded QPSK was employed as a modulation scheme. In addition, we assume that the BS is equipped with a one-dimension ULA with half-wavelength spacing and that the angle of arrival of the k-th user's ℓ-th MPC $\varphi_{k,\ell}$ is uniformly distributed in $[-\pi, \pi]$, resulting in $\varsigma_{k,\ell} = \sin(\varphi_{k,\ell})/2 \in [-1/2, 1/2]$.

5.5.1 Impact of Proposed Design

To validate the proposed design, we first investigated the relation between the phase transition and the empirical values of (ρ, δ), as shown in Fig. 5.5. It is seen from the figure that, when $K_{\mathrm{a}} = 50$, the (ρ, δ)s for $P = 24$ (2RBs) and $P = 36$ (3RBs) are located near the boundaries of $M = \infty$ and $M = 1$, respectively. These results suggest that, in the former and latter cases, the proposed GF-NOMA tends to fail and succeed, respectively, at accurately estimating the desired signals.

To clarify the impact of the proposed design and to confirm the aforementioned behavior, Fig. 5.6 shows the NMSE performance of the proposed scheme with SNR = 10 dB, $K_{\mathrm{a}} = 50$, and P varying from 24 to 36. Here, we define the NMSE as

$$\mathrm{NMSE} \triangleq \frac{\|\mathbf{H} - \hat{\mathbf{H}}\|_{\mathrm{F}}^2}{\|\mathbf{H}\|_{\mathrm{F}}^2}, \tag{5.53}$$

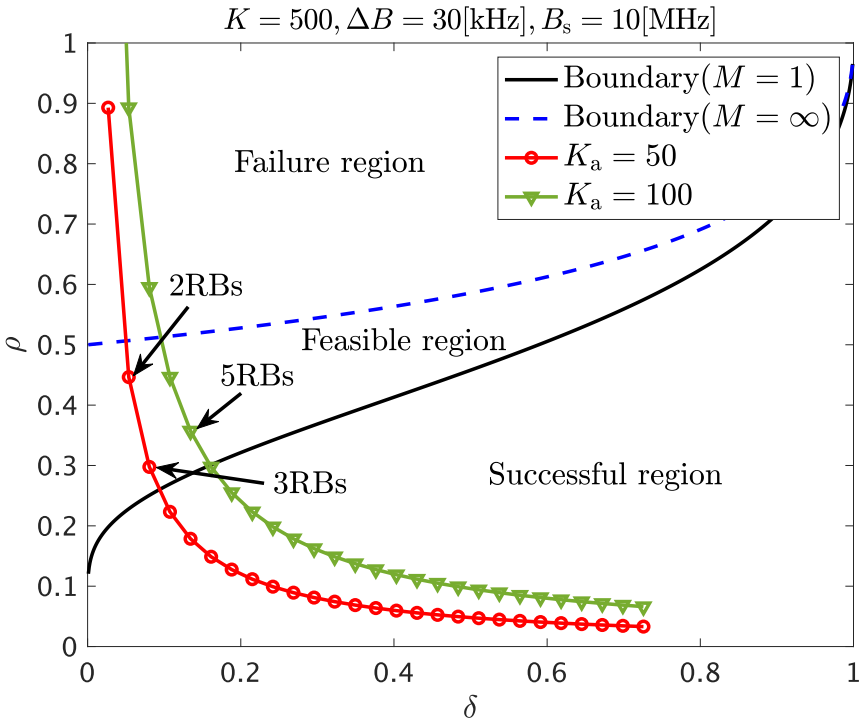

Fig. 5.5 Phase transitions in [106, 107] and empirical values of (ρ, δ).

where $\hat{\mathbf{H}}$ denotes the estimate of the CIR. Fig. 5.6 denotes the result produced by the LS method—which has knowledge of the non-zero rows in $\mathbf{H}$—by "Exact LS." It is obvious that the NMSE for $P = 24$ is significantly larger than it is for $P = 36$ and that the performance approaches the ideal (Exact LS) as P increases. The validity of the design for $K_a = 100$ active users is assessed in the next subsection.

5.5.2 NMSE Performance

We then compared the NMSE performance of the proposed scheme with that of a conventional scheme using the GMMV-AMP algorithm, *i.e.*, Scheme 1 of [64].[3] Fig. 5.7 and Fig. 5.8 show the NMSE performances for the $K_a = 50$ and $K_a = 100$ cases, respectively. The results indicate that the proposed phase-transition-based scheme significantly outperforms the conventional scheme and that the tailored signal model applied under the proposed scheme can produce a significantly higher gain than the DCS-theory based design in [64]. Furthermore, it is seen from Fig. 5.8 that

[3]For fair comparisons, we evaluated the case that the conventional scheme learns the hyperparameters corresponding to the sparsity ratio and the channel gains' variance by the EM algorithm, like our study.

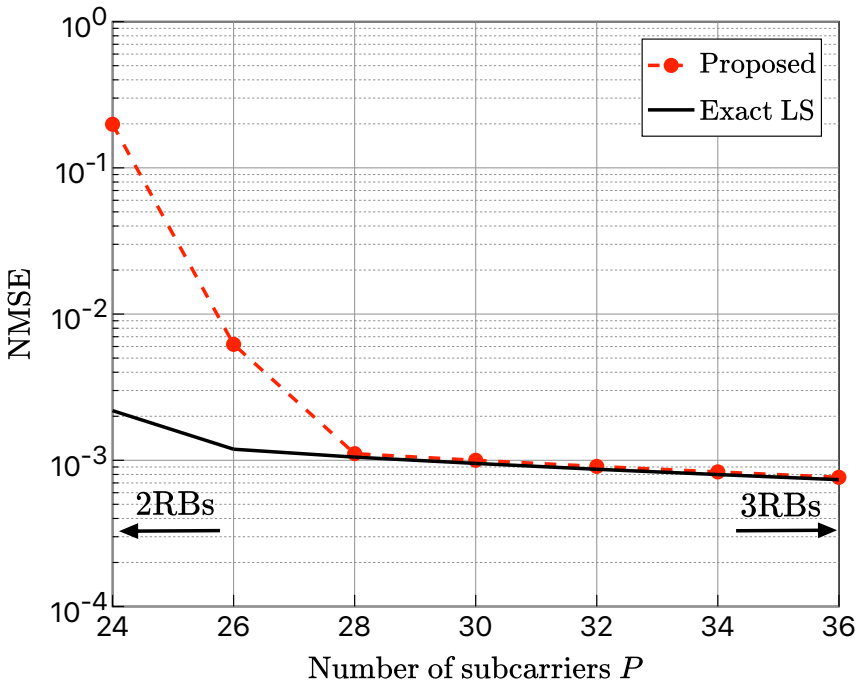

Fig. 5.6 NMSE performance as a function of P with SNR = 10 dB, $T = 28$, and $K_{\mathrm{a}} = 50$. The value of P varies from 24 (2RBs) to 36 (3RBs).

the proposed scheme achieves enhanced capability without lengthening the JACE time slots. This improvement can enable both massive system connectivity and low latency and represents the most notable advantage of the proposed scheme relative to conventional GF-NOMA schemes.

5.5.3 AUD Performance

In this section, we assess the AUD performance of the proposed method by evaluating its AER, which is defined by

$$\mathrm{AER} \triangleq \frac{\text{Number of miss-detected activity patterns}}{K}. \tag{5.54}$$

It is apparent that the AER is based on the probabilities of MD and FA. For the assessment of AER, we focused on a scenario similar to that in the related literature in which 10% of the users were active, namely $K_{\mathrm{a}} = 50$.

The AERs of the proposed and conventional schemes with $K_{\mathrm{a}} = 50$ and $P = 36$ (3RBs) are shown in Fig. 5.9. It is seen that, similar to the NMSE performance, the achievable AER of the proposed scheme is significantly lower than that of the

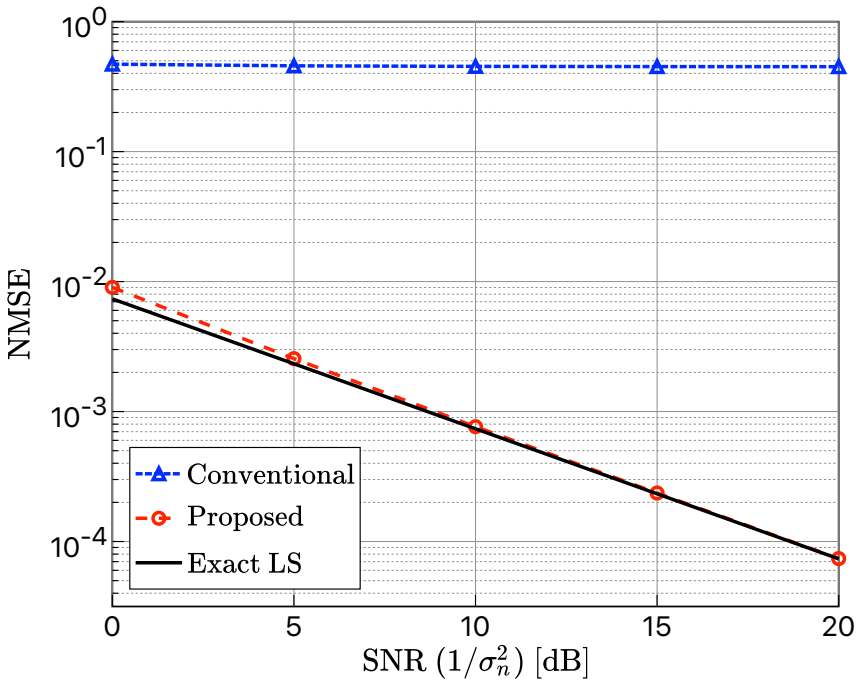

Fig. 5.7 NMSE performance of proposed and conventional schemes with $K_{\mathrm{a}} = 50$ active users, $T = 28$, and $P = 36$ (3RBs).

conventional scheme. Although an error floor for the proposed scheme has been induced by regarding an inactive user as an active user whose channel strength is quite small, the result indicates highly accurate AUD.

We further evaluated the AER performance with SNR = 10 dB, $K_{\mathrm{a}} = 50$, and the number of OFDM symbols varying from 14 to 56, as shown in Fig. 5.10, which also shows the performance results of the conventional scheme with $P = 60$ (5RBs). It is seen that, although the proposed scheme is inferior to the conventional scheme when $T < 24$ owing to the shrinkage of the dimensionality of the received signals, the former significantly outperforms the latter otherwise. Note that the performance of the proposed scheme can be improved by using more pilot subcarriers, even when T is small, as the dimensionality of the received signals is determined by the product of P and T. In contrast, it is seen that increasing the P in the conventional scheme does not yield any performance gain. These results indicate that the proposed GF-NOMA can take advantage of both the time and frequency domains to support massive users more efficiently than the conventional approach.

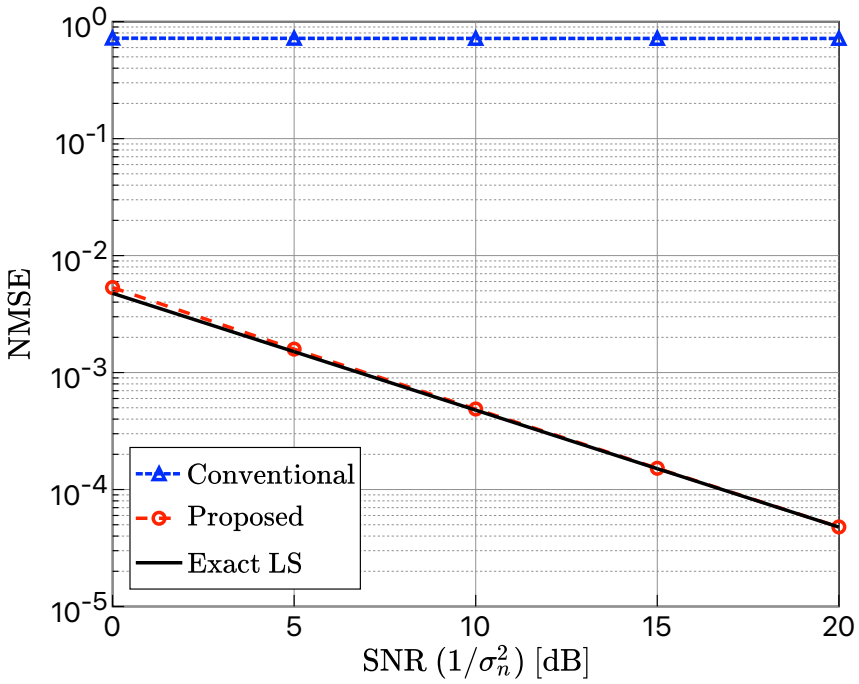

Fig. 5.8 NMSE performance of proposed and conventional schemes with $K_{\mathrm{a}} = 100$ active users, $T = 28$, and $P = 60$ (5RBs).

5.5.4 MUD Performance

The MUD performance of the proposed scheme was investigated in terms of its induced bit error rate (BER) and frame error rate (FER). As the accuracy of data detection is affected by the occurrence of MD, we took the number of lost bits and subframes owing to missing active users into account in defining the error rates as follows:

$$\mathrm{BER} \triangleq \frac{(K_{\mathrm{a}} - |\mathcal{A} \cap \hat{\mathcal{A}}|)T \log_2 Q + N_{b,e}}{K_{\mathrm{a}} T \log_2 Q}, \tag{5.55}$$

$$\mathrm{FER} \triangleq \frac{K_{\mathrm{a}} - |\mathcal{A} \cap \hat{\mathcal{A}}| + N_{f,e}}{K_{\mathrm{a}}}, \tag{5.56}$$

where $N_{b,e}$ and $N_{f,e}$ denote the number of bit and subframe errors, respectively, owing to the failure of data detection.

The BER and FER performance of the proposed scheme with $K_{\mathrm{a}} = 50$, $P = 36$, and $D = 18$ is shown in Fig. 5.11, in which the error rates of data detection using GaBP with Exact LS as well as those obtained using ideally performing perfect AUD and CE are also shown. The simulation results indicate that the proposed scheme can achieve performance approaching that of Exact LS even though it performs both

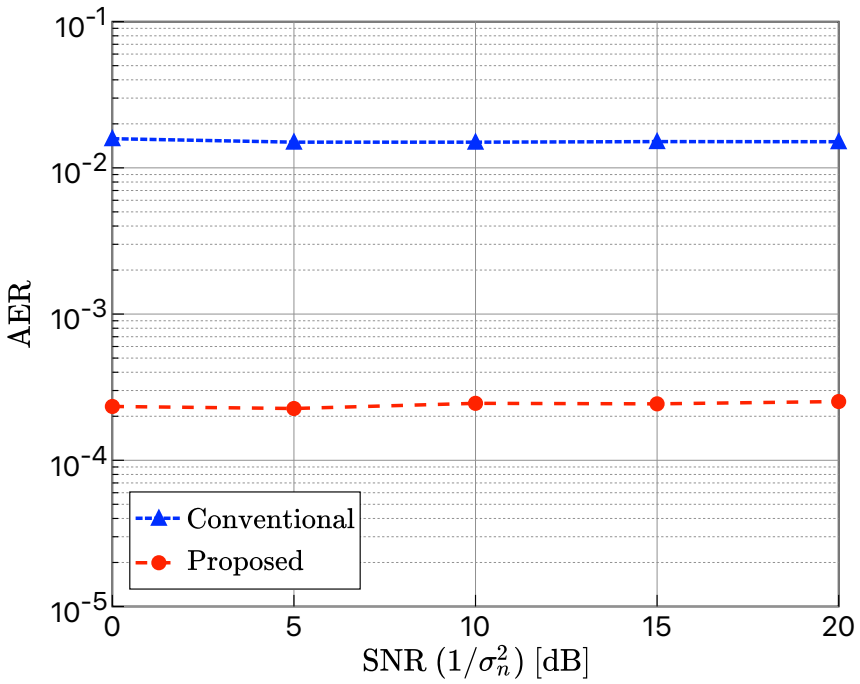

Fig. 5.9 AERs of proposed and conventional schemes with $K_{\mathrm{a}} = 50$, $T = 28$, and $P = 36$ (3RBs).

AUD and CE, with the performance gap decreasing as SNR increases, allowing the proposed method to converge on ideal performance. This convergence is induced by the performance degradation of GaBP arising from the correlation of CFRs, which is based on the relation in (5.40).

Fig. 5.12 shows the FER performance with $D = 16$ and 18 for $K_{\mathrm{a}} = 50$ and $P = 36$. For clarity, the proposed method is compared only with the GaBP having the ideal AUD and CE values as a benchmark. It is seen that there is an inevitable gap between the $D = 16$ and 18 for both the proposed scheme and the benchmark, indicating the necessity of adjusting D depending on the target reliability.

5.5.5 Effective Throughput

Finally, we evaluated the achievable throughput per active user of the proposed GF-NOMA in terms of the following effective throughput metric:

$$R_{\mathrm{eff}} \triangleq (1 - \mathrm{FER}) N_{\mathrm{sym}} T \log_2 Q, \tag{5.57}$$

where N_{sym} denotes the number of transmissible data symbols in a single OFDM symbol. Note that a system using the setup defined in Table 5.2 can use 24RBs for

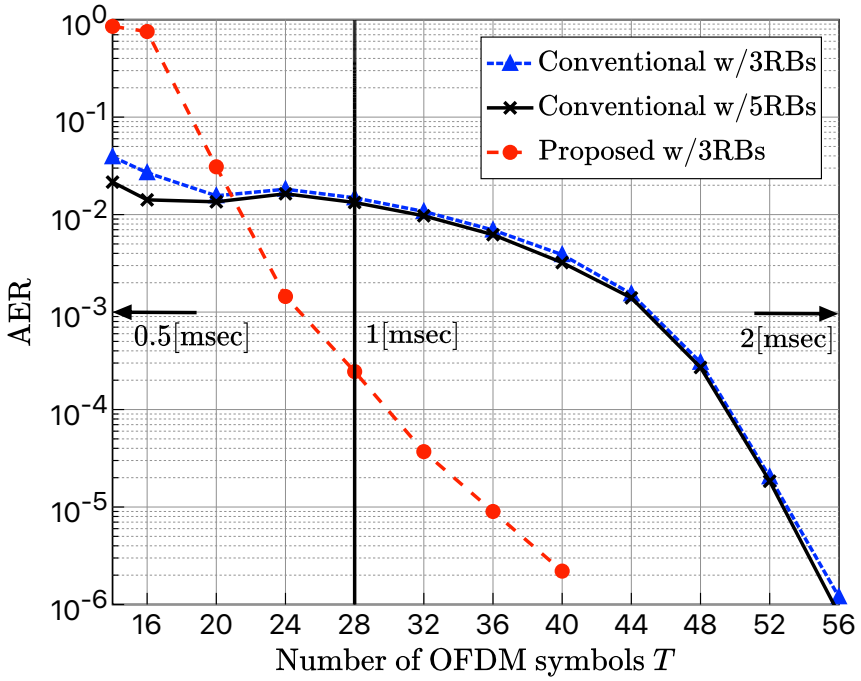

Fig. 5.10 AER as a function of T with SNR = 10 dB and $K_{\mathrm{a}} = 50$. The value of T varies from 14 (0.5 msec) to 56 (2 msec).

data transmission while supporting $K_{\mathrm{a}} = 50$ active users. Thus, when $D = 16$ and 18, each active user can transmit $N_{\mathrm{sym}} = 18$ and 16 data symbols in a single OFDM symbol, respectively.

Fig. 5.13 shows the effective throughput of the proposed scheme with $K_{\mathrm{a}} = 50$, $T = 28$, and $P = 36$. In addition to the benchmarks evaluated in Fig. 5.11, the figure plots the achievable maximum rate to indicate the performance limit of the proposed scheme. The results demonstrate that, while performing AUD, CE, and MUD, the proposed scheme enables active users to attain the maximum rate when the SNR is sufficiently high. It is further worth noting the high achievable throughput beyond 32[bytes/msec], indicating that the proposed GF-NOMA has the potential to satisfy a general requirement for URLLC defined under the 3GPP standard [111].[4] These results indicate that the proposed scheme is suitable for massive low-latency communication at moderate data rates.

Herein, we discuss the advantage of the proposed GF-NOMA over the K-repetition scheme [17]. For instance, under the setup of TABLE 5.2, each active user in the scheme with four repetitions has to use at least $16 \times 4 = 64$ subcarriers for data

[4]Concretely, the requirement is defined as $1 - 10^{-5}$ reliability within 1 msec user plane latency for 32 byte.

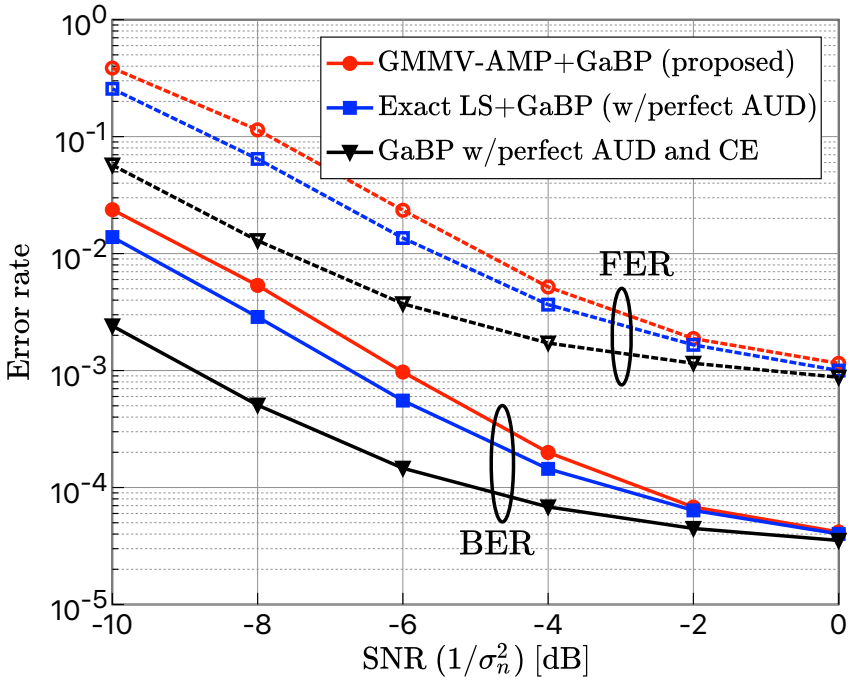

Fig. 5.11 BERs and FERs of the proposed scheme with $K_{\rm a} = 50$, $T = 28$, $P = 36$, and $D = 18$.

transmission to achieve 896 [bits/msec], which is the achievable maximum rate of the proposed GF-NOMA with $D = 18$. Then, although the system can divide the available subcarriers into $\lfloor 27 \times 12/64 \rfloor = 5$ groups, it has to support ten active users per group on average when $K_{\rm a} = 50$. As the probability that a replica of one active user is not collided is approximately $P_{\rm sc} = (4/5)^{49}$, the one that the K-repetition scheme with $K = 4$ can retrieve one transmitted packet is $1 - (1 - P_{\rm sc})^4 \approx 7.1 \times 10^{-5}$ even if each active user chooses the group for each replica transmission. This is quite lower than the decodable probability of the proposed GF-NOMA scheme. Moreover, even if the system adopts the SIC-based detection, the K-repetition scheme cannot relieve the issue fundamentally when the number of active users increases. Given the above, the proposed GF-NOMA scheme would be preferable to satisfy massive connectivity and low latency because of the efficient use of radio resources, *e.g.*, OFDM symbols, as compared with the K-repetition one.

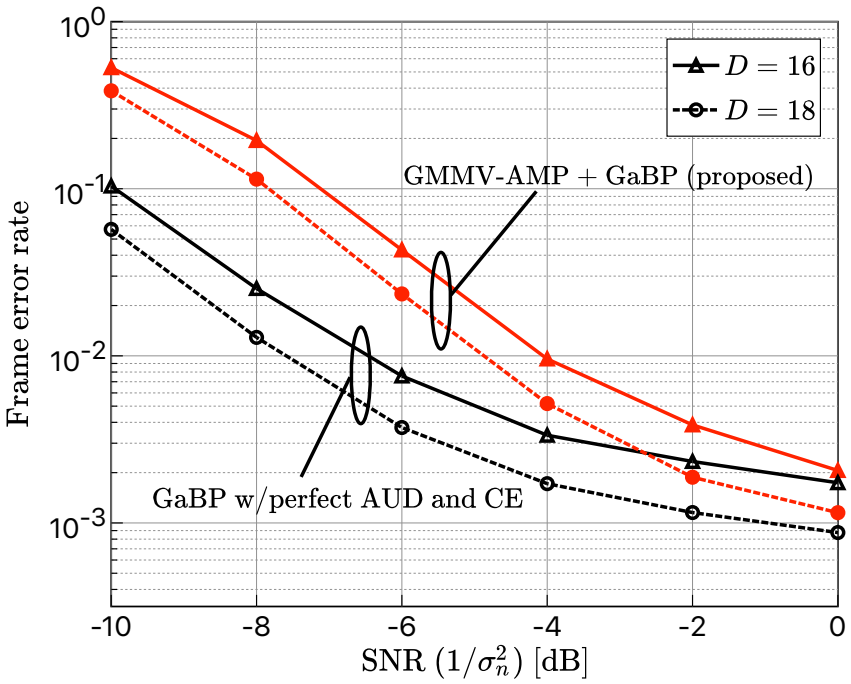

Fig. 5.12 FER of proposed scheme with $K_{\mathrm{a}} = 50$, $T = 28$, and $P = 36$.

5.6 Discussion of the Design of GF-NOMA

In this dissertation, we have considered three GF-NOMA systems using the spreading over 1) the time domain, 2) the frequency domain, and 3) both the time and frequency domains. In this section, we thus discuss each approach's advantageous region.

To begin with, we review three schemes considered in this dissertation. Chapters 2 and 3 have considered the GF-NOMA scheme using the spreading over the time domain, which is only applicable for a narrowband system under frequency-flat fading channels. On the other hand, Chapters 4 and 5 have considered an OFDM-based GF-NOMA system employing the spreading over the frequency domain, which can be applied to a wideband system under frequency-selective fading channels. It implies that the frequency selectivity of wireless channels is one of the crucial factors in the design of GF-NOMA. However, the OFDM-based GF-NOMA system can also employ the spreading over the time domain by using multiple OFDM symbols. Therefore, we here focus on the OFDM-based GF-NOMA system to discuss the advantageous region of three spreading patterns. Herein, as the scheme that uses the time domain, we consider the GF-NOMA scheme that uses a single subcarrier of multiple OFDM symbols to estimate active users and CFRs. This scheme is the special case of [64], where $P = 1$ and $T > 1$.

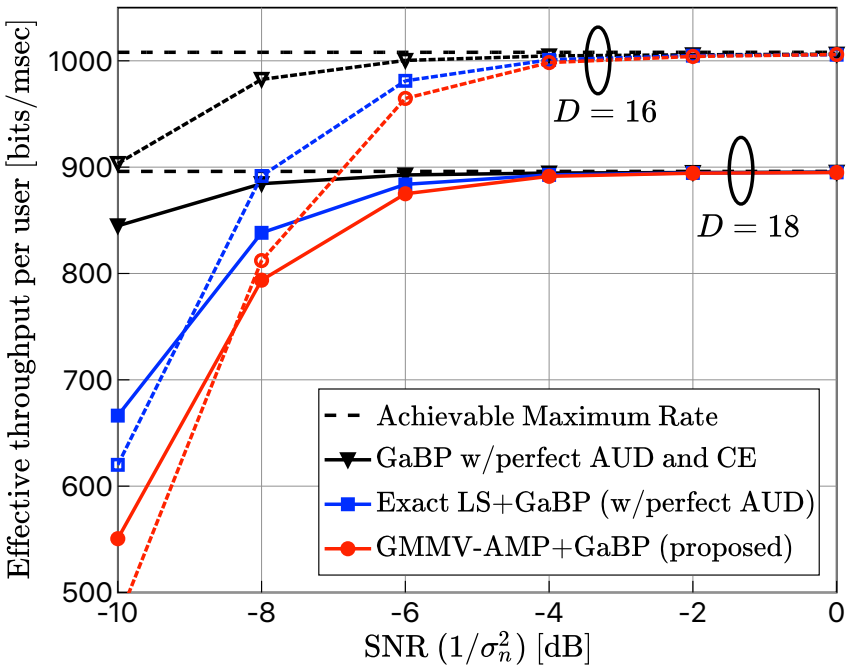

Fig. 5.13 Effective throughput of proposed scheme with $K_{\mathrm{a}} = 50$, $T = 28$, and $P = 36$.

To confirm the characteristics of three spreading patterns, TABLE 5.3 lists the computational complexity together with the values of ρ and δ in the phase transition analysis. Here, we consider the case that GMMV-AMP is utilized to perform AUD and CE. Notice that the performance of CS-based GF-NOMA schemes relies on the relation between ρ and δ, and ρ should be less than one.

Since the number of transmissible OFDM symbols, namely T, depends on the subcarrier spacing, the scheme spreading over the time domain can accommodate around a hundred active users under the latency requirements in the order of milliseconds if a wide subcarrier spacing such as 120 kHz is available at the system. This scheme is also applicable to massive IoT scenarios that do not impose stringent latency requirements. Moreover, we can see from TABLE 5.3 that its required computational complexity is lower than that of the other schemes.

In contrast, the scheme spreading over the frequency domain is suitable to the scenarios that the systems are demanded to use a narrow subcarrier spacing while meeting strict latency requirements, *e.g.*, less than 1 msec, since it uses only one OFDM symbol to perform AUD. However, as is evident from TABLE 5.3, it is difficult for this scheme to support many active users when L_{a} is large because it relies on the channel sparsity in the delay domain. This fact implies that the characteristic of CIRs has an impact on whether we need to use multiple OFDM symbols or not.

Table 5.3 Characteristics of the three spreading patterns.

Scheme	ρ	δ	Complexity order
Time domain only	K_{a}/T	T/K	$\mathcal{O}(KMT)$
Frequency domain only	$K_{\mathrm{a}}L_{\mathrm{path}}/P$	P/KL	$\mathcal{O}(KLMP)$
Time and frequency domains	$K_{\mathrm{a}}L_{\mathrm{path}}/PT$	PT/KL	$\mathcal{O}(KLMPT)$

Compared to the above schemes, the scheme exploiting both the time and frequency domains can accommodate many users under various latency requirements thanks to flexibly using radio resources. On the other hand, as shown in TABLE 5.3, its computational complexity would easily increase even if the GF-NOMA system employs a low-complexity message-passing algorithm such as GMMV-AMP.

Given the above, each of the three schemes has a different advantageous region, leading to the necessity of choosing the spreading pattern at the requirements for IoT applications.

5.7 Chapter Summary

In this chapter, we proposed a design for a new GF-NOMA for achieving both massive connectivity and low latency. Unlike three previous chapters, we introduced a tailored signal model for properly enlarging the dimensionality of measurements to enable highly accurate estimation while making full use of both the time and frequency domains. After that, we proposed a feasible system design, which takes the structure of 5G NR into account, based on the use of phase transitions by AMP algorithms. The proposed method further applies JACE via the GMMV-AMP algorithm to exploit both the sparsity arising from sporadic traffic and channel sparsity in the delay domain, enabling efficient MUD through the use of a symbol-wise GaBP algorithm. Numerical results indicate that the proposed scheme can outperform the state-of-the-art approach in handling low-latency massive-access scenarios. Moreover, we show the advantageous regions of the spreading patterns considered in this dissertation. We conclude that this chapter contributes to the establishment of GF-NOMA scheme that is suitable for massive access scenarios, especially massive low-latency communications.

Chapter 6

Conclusions and Future Work

Finally, this chapter concludes this dissertation and provides the future work on the study.

6.1 Conclusion

In this dissertation, we have proposed several approaches for the time-synchronous GF-NOMA systems and then addressed the possibility of realizing low latency communications by massive numbers of users, namely massive low-latency communications.

In Chapter 1, we first introduced the research background and the necessity of satisfying the requirements, namely *massive connectivity* and *low latency*. Then, we summarized the development of multiple access schemes, which motivated us to discuss grant-free multiple access schemes, especially GF-NOMA. After that, we described the fundamental challenges in GF-NOMA and reviewed the related works. Finally, we summarized the outline and contributions of this dissertation.

In Chapter 2, we investigated a narrowband GF-NOMA system with a multiple-antenna BS, where all users are time and frequency synchronized, and the BS can compensate for the (known) LSF coefficients of all users. We then proposed two low-complexity receivers based on MMV-AMP to perform AUD, CE, and MUD efficiently. We showed that our proposed receivers attain performance superior or comparable to that of the state-of-the-art BSASP while lowering the computational complexity. Remarkably, it was also shown that the proposed receiver exploiting MMV-AMP and the theoretically-redesigned threshold has the potential to be comparable to the code-domain grant-based NOMA in terms of SER performance.

Unlike Chapter 2, Chapter 3 examined GF-NOMA systems with spreading over the time-domain for two scenarios; 1) the BS does not know LSF coefficients of users, and

2) CFOs exist. For each scenario, we proposed alternative schemes and showed their superior performance via computer simulations. One of our proposals is the scheme to jointly perform AUD and CE via the EM-MMV-AMP algorithm that integrates the EM algorithm with the MMV-AMP algorithm. The other incorporates the transformation inspired by array-signal processing into the CD method, performing AUD in the presence of CFOs.

In contrast to two previous chapters, Chapter 4 investigated GF-NOMA systems employing the spreading pilot sequence over the frequency domain to further reduce the latency. Then, we proposed a hyperparameter-free receiver, which takes advantage of the sparsity of the channels in the delay domain. The proposed receiver is based on the CD method requiring no pre-tuning of hyperparameters to perform accurate estimation. We confirmed that the proposed scheme outperforms the conventional algorithms utilizing a block-sparsity or a sparsity of channels in the angular domain via computer simulations.

Finally, Chapter 5 proposed a design for a new GF-NOMA for achieving both massive connectivity and low latency. Unlike three previous chapters, we introduced a tailored signal model making full use of both the time and frequency domains, enabling the JACE that exploits both the sparsity arising from sporadic traffic and channel sparsity in the delay domain. After that, we proposed a feasible system design, which takes the structure of 5G NR into account, based on the use of phase transitions by AMP algorithms. Numerical results showed that the proposed scheme can outperform the state-of-the-art approach in handling low-latency massive-access scenarios. Moreover, we show the advantageous regions of the spreading patterns considered in this dissertation.

The methods discussed in this dissertation can meet the requirements mentioned earlier while overcoming the issues of the conventional schemes, leading to the enhancement of the applicability of GF-NOMA. For instance, the methods proposed in Chapters 2 and 3 have the potential to play a crucial role in supporting the NB-IoT applications. Moreover, time-sensitive applications such as motion control require real-time multi-device control. The proposed GF-NOMA schemes in Chapters 4 and 5 are applicable for such circumstances. In light of the above, we believe that the results in this dissertation will be the cornerstone of contributing to the development of future wireless communications systems for accommodating massive numbers of users.

6.2 Future Work

Although the proposals in this dissertation have contributed to satisfying both massive connectivity and low latency, there are several remaining topics for GF-NOMA schemes. In this section, we provide the future works on the study in this dissertation.

6.2.1 Design of GF-NOMA Schemes for Massive URLLC

Although the proposed scheme in Chapter 5 has contributed to the design of GF-NOMA systems to satisfy massive connectivity and low latency, it has not considered the reliability of data transmission. Moreover, the authors of [112] have recently considered the straightforward combination of the covariance-based AUD [50] with the 3GPP compliant pilot setting and found that URLLC requirements are still challenging from numerical results. Therefore, the realization of massive URLLC is still challenging, leading to the necessity of extensions of GF-NOMA schemes. A promising approach to this task is an extension of the scheme of Chapter 5, *e.g.*, a new design of GF-NOMA systems considering channel coding. Such a design is expected to realize future time-sensitive applications such as automated driving and factory automation.

6.2.2 OFDM-Based GF-NOMA in the Presence of CFOs

As shown in Fig. 1.3, as well as the related works, this dissertation did not investigate the OFDM-based GF-NOMA system in the presence of CFOs. On the other hand, as CFOs spoil the orthogonality among subcarriers of OFDM systems, we have to cope with it from a practical point of view.

To overcome the crucial issue, for multiuser MIMO-OFDM systems, some approaches to jointly estimate CFOs and channels have been investigated [113–116]. However, they require large numbers of antennas at the BS and sufficient training length to support more than ten users. It is obvious from this fact that even multiuser MIMO-OFDM systems, which does not AUD, have trouble in compensating CFOs. On the other hand, a frequency-asynchronous GF-NOMA system must estimate active users together with CFOs (and channels), further complicating its receiver design. Therefore, the design of OFDM-based GF-NOMA systems in the presence of CFOs will be a challenging research direction.

6.2.3 Design of Narrowband GF-NOMA Under Imperfect Time and Frequency Synchronization

As described in Section 6.1, this dissertation have considered the GF-NOMA systems under a perfect time synchronization; however, the precise timing control like grant-based systems is impractical due to the nature of grant-free transmission. In this context, since a narrowband system suffers from timing offsets, the receivers for an asynchronous narrowband GF-NOMA system have been recently investigated in [57–60].

On the other hand, as mentioned in Section 3.1, the systems comprised of low-cost devices, which are equipped with cheap crystal oscillators, suffer from CFOs [83]. If IoT applications utilize many low-cost devices, both timing offsets and CFOs should be compensated for. In light of the above, the receiver design for GF-NOMA systems under imperfect time and frequency synchronization is challenging and an interesting research direction. The solution to this task will contribute to massive IoT scenarios in which the network comprises of low-cost IoT devices.

6.2.4 Further Practical Receiver Design for GF-NOMA

In this dissertation, we have proposed several receivers for GF-NOMA systems. However, there is room for improvement in terms of practicality, as follows:

- Although Chapter 3 have proposed EM-MMV-AMP to perform AUD and CE without the knowledge of LSF coefficients, this scheme requires an activity ratio as prior information. An extension of EM-MMV-AMP, which overcomes the above issue, would be a promising solution to perform JACE in massive IoT scenarios, where it is difficult for the BS to obtain LSF coefficients.

- In addition to EM-MMV-AMP, in Chapter 3, we have proposed the AUD for a narrowband GF-NOMA system in the presence of CFOs. However, as shown in Section 3.4.3, the performance of this approach significantly degrades because of large-scale fading. From the above, extensions of the scheme are important to realize the systems, which do not require precise power control and frequency synchronization.

- Although a promising design of the GF-NOMA system has been proposed in Chapter 5, the required computational complexity would be easily high, as mentioned in Section 5.6. Therefore, further complexity reduction is needed to alleviate the issue.

Appendix A

Derivation of Equation (3.29)

The decision rule of (3.29) can be derived using certain approximations. First, according to (3.10), $\boldsymbol{\mu}_{t,k}$ can be approximated as

$$\boldsymbol{\mu}_{t,k} \approx \frac{\beta_{t,k}\tau_t}{\beta_{t,k}+\tau_t^2}\mathbf{v}_k,\ k \notin \mathcal{A}. \tag{A.1}$$

In addition, the squared norm of $\mathbf{v}_k$, *i.e.*, $\|\mathbf{v}_k\|_2^2$, is assumed to be M because $\mathbf{v}_k \sim \mathcal{CN}(\mathbf{v}_k; \mathbf{0}_M, \mathbf{I}_M)$. Thus, we can obtain the following relationship

$$\|\boldsymbol{\mu}_{t,k}\|_2^2 \approx M\left(\frac{\beta_{t,k}\tau_t}{\beta_{t,k}+\tau_t^2}\right)^2. \tag{A.2}$$

Hence, the update of $\beta_{t,k}$ for $k \notin \mathcal{A}$ satisfies the following inequality:

$$\begin{aligned}\beta_{t+1,n} &\approx \left(\frac{\beta_{t,k}\tau_t}{\beta_{t,k}+\tau_t^2}\right)^2 + \frac{\beta_{t,k}\tau_t^2}{\beta_{t,k}+\tau_t^2} \\ &= \frac{2\beta_{t,k}^2+\beta_{t,k}\tau_t^2}{\left(\beta_{t,k}+\tau_t^2\right)^2}\tau_t^2 \\ &< 2\tau_t^2 \quad \left(\because 2\beta_{t,k}^2+\beta_{t,k}\tau_t^2 < 2\left(\beta_{t,k}+\tau_t^2\right)^2\right).\end{aligned} \tag{A.3}$$

Based on the above, we can obtain the decision rule of (3.29).

References

[1] L. D. Xu, W. He, and S. Li, "Internet of things in industries: A survey," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233–2243, Nov. 2014.

[2] Cisco, "Cisco Annual Internet Report (2018-2023) White Paper," Mar. 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[3] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 55–61, Sep. 2017.

[4] S. Gangakhedkar, H. Cao, A. R. Ali, K. Ganesan, M. Gharba, and J. Eichinger, "Use cases, requirements and challenges of 5G communication for industrial automation," in *Proc. IEEE Int. Conf. Commun. Workshops*, Kansas City, MO, USA, May 2018, pp. 1–5.

[5] Y. Wu, X. Gao, S. Zhou, W. Yang, Y. Polyanskiy, and G. Caire, "Massive access for future wireless communication systems," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 148–156, Aug. 2020.

[6] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[7] N. H. Mahmood, H. Alves, O. A. López, M. Shehab, D. P. M. Osorio, and M. Latva-Aho, "Six key features of machine type communication in 6G," in *Proc. 2nd 6G Wireless Summit*, Levi, Finland, Mar. 2020, pp. 1–5.

[8] A. F. Molisch, *Wireless Communications.*, 2nd ed. Wiley Publishing, 2011.

[9] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge, U.K: Cambridge Univ. Press, 2005.

[10] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[11] A. Azari, P. Popovski, G. Miao, and C. Stefanovic, "Grant-free radio access for short-packet communications over 5G networks," in *Proc. IEEE Global Commun. Conf.*, Singapore, Dec. 2017, pp. 1–7.

[12] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 86–93, Jun. 2013.

[13] O. Y. Bursalioglu, C. Wang, H. Papadopoulos, and G. Caire, "RRH based massive MIMO with "on the fly" pilot contamination control," in *Proc. IEEE Int. Conf. Commun.*, Kuala Lumpur, Malaysia, May 2016, pp. 1–7.

[14] E. Björnson, E. de Carvalho, J. H. Sørensen, E. G. Larsson, and P. Popovski, "A random-access protocol for pilot allocation in crowded massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2220–2234, Apr. 2017.

[15] 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 Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sep. 2018.

[16] H. Chen *et al.*, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119–125, Dec. 2018.

[17] 3rd Generation Partnership Project. 3GPP, TR 38.214 Ver. 15.9.0, "5G; NR; Physical layer procedures for data," Apr. 2020.

[18] M. C. Lucas-Estañ, J. Gozalvez, and M. Sepulcre, "On the capacity of 5G NR grant-free scheduling with shared radio resources to support ultra-reliable and low-latency communications," *Sensors*, vol. 19, no. 16, p. 3575, Aug. 2019.

[19] N. Abramson, "The ALOHA system: Another alternative for computer communications," in *Proc. Fall Joint Computer Conf.*, New York, NY, Nov. 1970, pp. 281–285.

[20] L. G. Roberts, "Aloha packet system with and without slots and capture," *SIGCOMM Comput. Commun. Rev.*, vol. 5, no. 2, pp. 28–42, Apr. 1975.

[21] E. Casini, R. De Gaudenzi, and O. Herrero, "Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1408–1419, Apr. 2007.

[22] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 477–487, Feb. 2011.

[23] E. Paolini, G. Liva, and M. Chiani, "Coded slotted ALOHA: A graph-based method for uncoordinated multiple access," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6815–6832, Dec. 2015.

[24] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[25] C. Stefanović, P. Popovski, and D. Vukobratovic, "Frameless ALOHA protocol for wireless networks," *IEEE Commun. Lett.*, vol. 16, no. 12, pp. 2087–2090, Dec. 2012.

[26] R. Ahlswede, "Multi-way communication channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Sep. 1971, pp. 23–52.

[27] H. Liao, "A coding theorem for multiple access communications," in *Proc. IEEE Int. Symp. Inf. Theory*, Pacific Grove, CA, USA, 1972.

[28] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, Jan. 1985.

[29] X. Chen and D. Guo, "Gaussian many-access channels: Definition and symmetric capacity," in *Proc. IEEE Inf. Theory Workshops*, Seville, Spain, Sep. 2013, pp. 1–5.

[30] X. Chen, T.-Y. Chen, and D. Guo, "Capacity of Gaussian many-access channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3516–3539, Jun. 2017.

[31] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 2523–2527.

[32] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 3rd Quart. 2020.

[33] F. Wei, Y. Wu, W. Chen, W. Yang, and G. Caire, "On the fundamental limits of MIMO massive multiple access channels," in *Proc. IEEE Int. Conf. Commun.*, Shanghai, China, May 2019, pp. 1–6.

[34] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 3rd Quart. 2018.

[35] R.-A. Stoica, G. T. F. de Abreu, T. Hara, and K. Ishibashi, "Massively concurrent non-orthogonal multiple access for 5G networks and beyond," *IEEE Access*, vol. 7, pp. 82 080–82 100, 2019.

[36] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. 77th IEEE Veh. Technol. Commun.*, Dresden, Germany, Jun. 2013, pp. 1–5.

[37] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-s. Kwak, "Power-domain non-orthogonal multiple sccess (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart. 2017.

[38] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient random access for the quasi-static fading MAC," in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, Jul. 2019, pp. 2768–2772.

[39] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Grant-free massive random access with a massive MIMO receiver," in *Proc. 53rd Asilomar Conf. Signals, Syst., and Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 23–30.

[40] V. K. Amalladinne, A. Vem, D. K. Soma, K. R. Narayanan, and J. Chamberland, "A coupled compressive sensing scheme for unsourced multiple access," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, Calgary, AB, Canada, Apr. 2018, pp. 6628–6632.

[41] A. Fengler, P. Jung, and G. Caire, "SPARCs ans AMP for unsourced random access," in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, Jul. 2019, pp. 2843–2847.

[42] A. Vem, K. R. Narayanan, J. Chamberland, and J. Cheng, "A user-independent successive interference cancellation based coding scheme for the unsourced random access Gaussian channel," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8258–8272, Dec. 2019.

[43] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan, and J.-F. Chamberland, "Polar coding and random spreading for unsourced multiple access," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, Jun. 2020, pp. 1–6.

[44] A. Decurninge, I. Land, and M. Guillaud, "Tensor-based modulation for unsourced massive random access," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 552–556, Mar. 2021.

[45] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[46] T. Ding, X. Yuan, and S. C. Liew, "Sparsity learning-based multiuser detection in grant-free massive-device multiple access," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3569–3582, Jul. 2019.

[47] Q. Zou, H. Zhang, D. Cai, and H. Yang, "A low-complexity joint user activity, channel and data estimation for grant-free massive MIMO systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 1290–1294, 2020.

[48] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, and W. Yu, "Grant-free access via bilinear inference for cell-free MIMO with low-coherence pilots," *IEEE Trans. Wireless Commun.*, pp. 7694–7710, Nov. 2021.

[49] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

[50] S. Haghighatshoar, P. Jung, and G. Caire, "Improved scaling law for activity detection in massive MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Vail, CO, USA, Jun. 2018, pp. 381–385.

[51] Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Covariance based joint activity and data detection for massive random access with massive MIMO," in *Proc. IEEE Int. Conf. Commun.*, Shanghai, China, May 2019, pp. 1–6.

[52] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2925–2951, May 2021.

[53] Y. Li, M. Xia, and Y.-C. Wu, "Activity detection for massive connectivity under frequency offsets via first-order algorithms," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1988–2002, Mar. 2019.

[54] L. Liu and W. Yu, "Massive connectivity with massive MIMO–Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.

[55] T. Jiang, Y. Shi, J. Zhang, and K. B. Letaief, "Joint activity detection and channel estimation for IoT networks: Phase transition and computation-estimation tradeoff," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6212–6225, Aug. 2019.

[56] X. Shao, X. Chen, and R. Jia, "A dimension reduction-based joint activity detection and channel estimation algorithm for massive access," *IEEE Trans. Signal Process.*, vol. 68, pp. 420–435, Dec. 2019.

[57] J. Fu, G. Wu, Y. Zhang, L. Deng, and S. Fang, "Active user identification based on asynchronous sparse Bayesian learning with SVM," *IEEE Access*, vol. 7, pp. 108 116–108 124, 2019.

[58] C.-T. Liu, H.-J. Su, and Y. Takano, "Sparse activity, timing detection and channel estimation for grant-free uplink communications," in *Proc. IEEE 31th Int. Symp. Pers., Indoor and Mobile Radio Commun.*, London, UK, Aug.-Sep. 2020, pp. 1–7.

[59] L. Liu and Y. F. Liu, "An efficient algorithm for device detection and channel estimation in asynchronous iot systems," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, Toronto, ON, Canada, Jun. 2021, pp. 4815–4819.

[60] R. B. D. Renna and R. C. de Lamare, "Dynamic message scheduling based on activity-aware residual belief propagation for asynchronous mMTC," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1290–1294, Jun. 2021.

[61] N. Y. Yu, K. Lee, and J. Choi, "Pilot signal design for compressive sensing based random access in machine-type communications," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, San Francisco, CA, USA, Mar. 2017, pp. 1–6.

[62] X. Shao, X. Chen, C. Zhong, and Z. Zhang, "Joint activity detection and channel estimation for mmW/THz wideband massive access," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2020, pp. 1–6.

[63] M. Ke, Z. Gao, Y. Wu, and X. Meng, "Compressive massive random access for massive machine-type communications (mMTC)," in *Proc. IEEE Global Conf. Signal and Inf. Process.*, Anaheim, CA, USA, Nov. 2018, pp. 156–160.

[64] 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 Trans. Signal Process.*, vol. 68, pp. 764–779, Jan. 2020.

[65] 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 Commun. Lett.*, vol. 7, no. 4, pp. 682–685, Aug. 2018.

[66] A. Bayesteh, E. Yi, H. Nikopour, and H. Baligh, "Blind detection of SCMA for uplink grant-free multiple-access," in *Proc. 11th Int. Symp. Wireless Commun. Syst.*, Barcelona, Spain, Aug. 2014, pp. 853–857.

[67] G. Hannak, M. Mayer, A. Jung, G. Matz, and N. Goertz, "Joint channel estimation and activity detection for multiuser communication systems," in *Proc. IEEE Int. Conf. Commun. Workshop*, London, UK, Jun. 2015, pp. 2086–2091.

[68] F. Wei, W. Chen, Y. Wu, J. Ma, and T. A. Tsiftsis, "Message-passing receiver design for joint channel estimation and data decoding in uplink grant-free SCMA systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 167–181, Jan. 2019.

[69] S. Jiang, X. Yuan, X. Wang, C. Xu, and W. Yu, "Joint user identification, channel estimation, and signal detection for grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6960–6976, Oct. 2020.

[70] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing–Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014.

[71] ——, "Bilinear generalized approximate message passing–Part II: Applications," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5854–5867, Nov. 2014.

[72] S. Rangan, A. K. Fletcher, V. K. Goyal, E. Byrne, and P. Schniter, "Hybrid approximate message passing," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4577–4592, Sep. 2017.

[73] K. Ito, T. Takahashi, S. Ibi, and S. Sampei, "Bilinear Gaussian belief propagation for large MIMO channel and data estimation," in *Proc. IEEE Global Commun. Conf.*, Taipei, Taiwan, Dec. 2020, pp. 1–6.

[74] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.

[75] A. Hjrungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications.* New York, NY, USA: Cambridge University Press, 2011.

[76] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18 914–18 918, Nov. 2009.

[77] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborová, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Brisbane, QLD, Australia, Apr. 2015, pp. 2021–2025.

[78] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul.-Aug. 2011, pp. 2168–2172.

[79] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[80] J. Kim, W. Chang, B. Jung, D. Baron, and J. C. Ye, "Belief propagation for joint sparse recovery," Feb. 2011. [Online]. Available: https://arxiv.org/abs/1102.3289.

[81] J. Bausch, "On the efficient calculation of a linear combination of chi-square random variables with an application in counting string vacua," *J. Phys. A Math. and Theor.*, vol. 46, no. 50, Nov. 2013.

[82] A. M. Mathai and R. K. Saxena, *The H-function with Applications in Statistics and Other Disciplines.* New York, NY, USA: Wiley, 1978.

[83] J. Xu, J. Yao, L. Wang, Z. Ming, K. Wu, and L. Chen, "Narrowband Internet of Things: Evolutions, technologies, and open issues," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1449–1462, Jun. 2018.

[84] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[85] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[86] M. Dudek, I. Nasr, G. Bozsik, M. Hamouda, D. Kissinger, and G. Fischer, "System analysis of a phased-array radar applying adaptive beam-control for future automotive safety applications," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 34–47, Jan. 2015.

[87] R. M. Gray, *Toeplitz and Circulant Matrices: A Review.* Boston, MA, USA: Now Publishers, 2006.

[88] E. J. Candés, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, Aug. 2013.

[89] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Prob.*, vol. 31, no. 11, p. 115002, 2015.

[90] N. J. Myers and R. W. Heath Jr., "Joint CFO and channel estimation in millimeter wave systems with one-bit ADCs," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, Curacao, Dec. 2017, pp. 1–5.

[91] ——, "Message passing-based joint CFO and channel estimation in mmWave systems with one-bit ADCs," *IEEE Trans. Signal Process.*, vol. 18, no. 6, pp. 3064–3077, Jun. 2019.

[92] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Statist.*, vol. 21, no. 1, pp. 124–127, 1950.

[93] C. Rusu, N. G. Prelcic, and R. W. H. Jr, "Algorithms for the construction of incoherent frames under various design constraints," *Signal Process.*, vol. 152, pp. 363–372, Nov. 2018.

[94] X. Wu, L. Gu, W. Wang, and X. Gao, "Pilot design and AMP-based channel estimation for massive MIMO-OFDM uplink transmission," in *Proc. IEEE 27th Int. Symp. Pers., Indoor and Mobile Radio Commun.*, Valencia, Spain, Sep. 2016, pp. 1–7.

[95] C. Steffens, M. Pesavento, and M. E. Pfetsch, "A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1483–1497, Mar. 2018.

[96] Y. Zhou, M. herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statics and capacity via the virtual channel representation," Univ. Wisconsin-Madison Tech. Rep., Feb. 2007.

[97] L.-L. Yang, *Multicarrier Communications.* Chichester, U.K: John Wiley, 2009.

[98] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, "Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1356–1368, Mar. 2015.

[99] S. Li, G. Zhao, W. Zhang, Q. Qiu, and H. Sun, "ISAR imaging by two-dimensional convex optimization-based compressive sensing," *IEEE Sensors J.*, vol. 16, no. 19, pp. 5437–5451, Oct. 2016.

[100] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, Nov. 2010.

[101] T. Takahashi, S. Ibi, and S. Sampei, "Design of adaptively scaled belief in large MIMO detection for higher-order modulation," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, Kuala Lumpur, Malaysia, Dec. 2017, pp. 1800–1805.

[102] ——, "Design of adaptively scaled belief in multi-dimensional signal detection for higher-order modulation," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1986–2001, Mar. 2019.

[103] 3rd Generation Partnership Project. 3GPP, TR 25.996 Ver. 14.2.0, "Study on new radio (NR) access technology; Physical layer aspects," Sep. 2017.

[104] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *Proc. IEEE Inf. Theory Workshop*, Cairo, Egypt, Jan. 2010, pp. 1–6.

[105] M. Simon, "On the probability density function of the squared envelope of a sum of random phase vectors," *IEEE Trans. Commun.*, vol. 33, no. 9, pp. 993–996, Sep. 1985.

[106] D. L. Donoho, I. Jhonstone, and A. Montanari, "Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3396–3433, Jun. 2013.

[107] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)," *IEEE Trans. Inf. Theory*, vol. 7, no. 59, pp. 4290–4308, Jul. 2013.

[108] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, Jan. 2017.

[109] M. Pretti, "A message-passing algorithm with damping," *J. Stat. Mech., Theory Exp.*, vol. 2005, p. 11008, Nov. 2005.

[110] P. Som, T. Datta, A. Chockalingam, and B. S. Rajan, "Improved large-MIMO detection based on damped belief propagation," in *Proc. IEEE Inf. Theory Workshop*, Cairo, Egypt, Jan. 2010, pp. 1–5.

[111] 3rd Generation Partnership Project. 3GPP, TR 38.913 v16.0.0, "Study on scenarios and requirements for next generation access technologies," Jul. 2020.

[112] H. Yan, A. Ashikhmin, and H. Yang, "Can massive MIMO support URLLC?" in *Proc. IEEE 93rd Veh. Technol. Conf.*, Helsinki, Finland, Apr. 2021, pp. 1–5.

[113] W. Zhang and F. Gao, "Blind frequency synchronization for multiuser OFDM uplink with large number of receive antennas," *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2255–2268, May 2016.

[114] W. Zhang, F. Gao, S. Jin, and H. Lin, "Frequency synchronization for uplink massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 235–249, Jan. 2018.

[115] Y. Feng, W. Zhang, F. Gao, and Q. Sun, "Computationally efficient blind CFO estimation for massive MIMO uplink," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7795–7799, Aug. 2018.

[116] Z. Mokhtari, M. Sabbaghian, and T. Eriksson, "Iterative channel and CFO estimation for SC-FDE and OFDM based massive MIMO systems," in *Proc. IEEE 89th Veh. Technol. Commun.*, Kuala Lumpur, Malaysia, Apr.-May 2019, pp. 1–5.

Publications

Related Journal Papers

1. **Takanori Hara**, Hiroki Iimori, and Koji Ishibashi, "Hyperparameter-free receiver for grant-free NOMA systems with MIMO-OFDM," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 810–814, Apr. 2021. (Copyright© 2021 IEEE) (Chapter 4)

2. **Takanori Hara** and Koji Ishibashi, "Grant-free non-orthogonal multiple access with multiple-antenna base station and its efficient receiver design," *IEEE Access*, vol. 7, pp. 175717–175726, Nov. 2019. (Copyright© 2019 IEEE) (Chapter 2)

Related Conference Papers

1. **Takanori Hara**, Hiroki Iimori, and Koji Ishibashi, "Activity detection for uplink grant-free NOMA in the presence of carrier frequency offsets," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops 2020)*, Dublin Ireland Jun. 2020. (Copyright© 2020 IEEE) (Chapter 3)

2. **Takanori Hara** and Koji Ishibashi, "Grant-free NOMA using approximate message passing with multi-measurement vector," in *Proc. 34th Int. Conf. Inf. Networking (ICOIN 2020)*, Barcelona, Spain, Jan. 2020. (Copyright© 2020 IEEE) (Chapter 2)

3. **Takanori Hara** and Koji Ishibashi, "Low complexity uplink grant-free NOMA based on boosted approximate message passing," in *Proc. 53rd Asilomar Conf. Signals, Syst., and Comput.*, Pacific Grove, CA, USA, Nov. 2019. (Copyright© 2019 IEEE) (Chapter 2)

Other Journal Papers

1. **Takanori Hara** and Koji Ishibashi, "Blind multiple measurement vector AMP based on expectation maximization for grant-free NOMA," *IEEE Wireless Commun. Lett.* (under review)

2. **Takanori Hara**, Ryuhei Takahashi, and Koji Ishibashi, "Ambient OFDM pilot-aided backscatter communications: Concept and design," *IEEE Access*, vol. 9, pp. 89210–89221, Jun. 2021.

3. Hiroki Iimori, Giussepe Thadeu Freitas de Abreu, Omid Taghizadeh, Razvan-Andrei Stoica, **Takanori Hara**, and Koji Ishibashi, "A stochastic gradient descent approach for hybrid MmWave beamforming with blockage and CSI-error robustness," *IEEE Access*, vol. 9, pp. 74471–74487, May 2021.

4. Hiroki Iimori, Giussepe Thadeu Freitas de Abreu, **Takanori Hara**, Koji Ishibashi, Razvan-Andrei Stoica, David González G., and Osvaldo Gonsa, "Robust symbol detection in large-scale overloaded NOMA systems," *IEEE Open J. Commun. Society*, vol. 2, pp. 512–533, Mar. 2021.

5. Hiroki Iimori, Giussepe Thadeu Freitas de Abreu, Omid Taghizadeh, Razvan-Andrei Stoica, **Takanori Hara**, and Koji Ishibashi, "Stochastic learning robust beamforming for millimeter-wave systems with path blockage," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1557–1561, Sep. 2020.

6. Razvan-Andrei Stoica, Giussepe Thadeu Freitas de Abreu, **Takanori Hara**, and Koji Ishibashi, "Massively concurrent non-orthogonal multiple access for 5G networks and beyond," *IEEE Access*, vol. 7, pp. 82080–82100, Jun. 2019.

Other Conference Papers

1. Ryota Yatsu, **Takanori Hara**, Koji Ishibashi, Sota Tsuchiya, and Hideki Endo, "Packet aggregation based on encryption-then-compression for highly efficient multi-hop transmission," in *Proc. Asia-Pacific Signal and Inf. Process. Assoc. Annu. Summit and Conf.*, Auckland, New Zealand, Dec. 2020.

2. Takahide Murakami, Hiroyuki Shinbo, Yu Tsukamoto, Shinobu Nanba, Yoji Kishi, Morihiko Tamai, Hiroyuki Yokoyama, **Takanori Hara**, Koji Ishibashi, Kensuke Tsuda, Yoshimi Fujii, Fumiyuki Adachi, Keisuke Kasai, Masataka Nakazawa,

Yuta Seki, and Takayuki Sotoyama, "Research project to realize various high-reliability communications in advanced 5G network," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, Seoul, Korea (South), May 2020.

3. **Takanori Hara**, Razvan-Andrei Stoica, Koji Ishibashi, and Giussepe Thadeu Freitas de Abreu, "On the sum-rate capacity and spectral efficiency gains of massively concurrent NOMA systems," in *Proc. IEEE Wireless Commun. and Netw. Conf.*, Marrakech, Morocco, Apr. 2019.

4. **Takanori Hara**, Koji Ishibashi, Soon Xin Ng, and Lajos Hanzo, "Low-complexity generator polynomial search for turbo trellis-coded spatial modulation using symbol-based EXIT charts," in *Proc. IEEE 10th Int. Symp. Turbo Codes and Iterative Inf. Process.*, Hong Kong, China, Dec. 2018.

5. Kazuya Ohira, **Takanori Hara**, and Koji Ishibashi, "Aggregate-compression-aided subcarrier IQ index modulation," in *Proc. IEEE Workshop Positioning, Netw., and Commun.*, Bremen, Germany, Oct. 2018.

Domestic Conference Papers

1. **Takanori Hara** and Koji Ishibashi, "Multiple measurement vector approximate message passing without prior information of channels for grant-free non-orthogonal multiple access," IEICE Tech. Rep., vol. 121, no. 72, RCS2021-71, pp. 243–248, Jun. 2021.

2. **Takanori Hara** and Koji Ishibashi, "Grant-free non-orthogonal transmission for low-latency and massive connectivity," IEICE General Conference 2021, B-5-70, Mar. 2021.

3. Ryo Sugawara, **Takanori Hara**, and Koji Ishibashi, "A study on iterative decoding of polarization-adjusted convolutional codes," IEICE General Conference 2021, A-2-3, Mar. 2021.

4. **Takanori Hara** and Koji Ishibashi, "Massive grant-free non-orthogonal multiple access utilizing time and frequency spreading," IEICE Tech. Rep., vol. 120, no. 298, RCS2020-134, pp. 1–6, Dec. 2020.

5. **Takanori Hara**, Hiroki Iimori, and Koji Ishiabshi, "Hyperparameter-free receiver for OFDM-based grant-free non-orthogonal multiple access," IEICE Society Conference 2020, B-5-11, Sep. 2020.

6. **Takanori Hara** and Koji Ishibashi, "A study on non-orthogonal pilot design for grant-free multi-user massive MIMO," IEICE Tech. Rep., vol. 120, no. 29, RCS2020-22, pp. 67–72, May 2020.

7. **Takanori Hara**, Ryuhei Takahashi, and Koji Ishibashi, "Generalization of performance analysis on OFDM pilot-aided ambient backscatter communications," IEICE Tech. Rep., vol. 120, no. 10, RCS2020-9, pp. 49–54, Apr. 2020.

8. Takahide Murakami, Hiroyuki Shinbo, Yu Tsukamoto, Shinobu Nanba, Yoji Kishi, Morihiko Tamai, Hiroyuki Yokoyama, **Takanori Hara**, Koji Ishibashi, Kensuke Tsuda, Yoshimi Fujii, Fumiyuki Adachi, Keisuke Kasai, Masataka Nakazawa, Yuta Seki, and Takayuki Sotoyama, "R&D of technology for high reliability management of advanced 5G network to various requirements of communication services," IEICE Tech. Rep., vol. 119, no. 448, RCS2019-320, pp. 1–6, Mar. 2020.

9. **Takanori Hara** and Koji Ishibashi, "Transmission efficiency of uplink grant-free non-orthogonal multiple access," IEICE Tech. Rep., vol. 119, no. 448, RCS2019-325, pp. 31–36, Mar. 2020.

10. **Takanori Hara** and Koji Ishibashi, "Active user detection for uplink grant-free non-orthogonal multiple access in the presence of carrier frequency offsets," IEICE Tech. Rep., vol. 119, no. 378, RCS2019-319, pp. 313–317, Jan. 2020.

11. **Takanori Hara** and Koji Ishibashi, "A study on active user detection and channel estimation for grant-free non-orthogonal multiple access based on OFDM," IEICE Tech. Rep., vol. 119, no. 345, RCS2019-237, pp. 1–5, Dec. 2019.

12. Ryuhei Takahashi, **Takanori Hara**, and Koji Ishibashi, "OFDM pilot-aided delay-shift keying and its optimal detection for ultra-low power communications," The 42nd Symposium on Information Theory and its Applications (SITA2019), pp. 423–428, Nov. 2019.

13. **Takanori Hara** and Koji Ishibashi, "Low-complexity grant-free non-orthogonal multiple access based on approximate message passing," The 42nd Symposium on Information Theory and its Applications (SITA2019), pp. 323–328, Nov. 2019.

14. **Takanori Hara** and Koji Ishibashi, "A study on low-complexity data detection for grant-free non-orthogonal multiple access," IEICE Tech. Rep., vol. 119, no. 296, RCS2019-211, pp. 49–54, Nov. 2019.

15. **Takanori Hara** and Koji Ishibashi, "Theoretical analysis on grant-free non-orthogonal multiple access using vector approximate message passing," IEICE Tech. Rep., vol. 119, no. 244, RCS2019-179, pp. 7–12, Oct. 2019.

16. **Takanori Hara** and Koji Ishibashi, "Generalized boosted approximate message passing for uplink grant-free non-orthogonal multiple access," IEICE Tech. Rep., vol. 119, no. 176, RCS2019-148, pp. 13–18, Aug. 2019.

17. **Takanori Hara** and Koji Ishibashi, "Boosted approximate message passing for uplink grant-free non-orthogonal multiple access," IEICE Tech. Rep., vol. 119, no. 109, SR2019-26, pp.43–48, Jul. 2019.

18. **Takanori Hara** and Koji Ishibashi, "Channel estimation and data Detection for grant-free non-orthogonal multiple access using vector approximate message passing," IEICE Tech. Rep., vol. 119, no. 90, RCS2019-96, pp. 347–352, Jun. 2019.

19. **Takanori Hara**, Razvan-Andrei Stoica, Giussepe Abreu, and Koji Ishibashi, "Low-complexity detection via Gaussian belief propagation for frame-theoretic non-orthogonal multiple access," The 41st Symposium on Information Theory and its Applications (SITA2018), pp. 188–193, Dec. 2018.

20. Kazuya Ohira, **Takanori Hara**, and Koji Ishibashi, "Aggregate-compression-aided subcarrier IQ index modulation with discreteness-aware approximate message passing," IEICE Tech. Rep., vol. 118, no. 125, RCS2018-128, pp. 243–248, Jul. 2018.

21. **Takanori Hara**, Koji Ishibashi, and Lajos Hanzo "Low-complexity code search method for turbo trellis-coded spatial modulation using symbol-based EXIT charts," IEICE Tech. Rep., vol. 118, no. 101, RCS2018-70, pp. 207–212, Jun. 2018.

22. **Takanori Hara** and Koji Ishibashi, "A study on lattice based on construction A with nonbinary repeat-accumulate codes," IEICE General Conference 2017, A-2-12, Mar. 2017.