# NOVEL UNSOURCED RANDOM ACCESS ALGORITHMS OVER GAUSSIAN AND FADING CHANNELS

A DISSERTATION SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ELECTRICAL AND ELECTRONICS ENGINEERING

By

Mohammad Javad Ahmadi

January 2024

Novel Unsourced Random Access Algorithms over Gaussian and Fading Channels
By Mohammad Javad Ahmadi
January 2024

We certify that we have read this dissertation and that in our opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Tolga Mete Duman (Advisor)

Ezhan Karaşan

Sinan Gezici

Tolga Girici

Ayşe Melda Yüksel Turgut

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

# ABSTRACT

# NOVEL UNSOURCED RANDOM ACCESS ALGORITHMS OVER GAUSSIAN AND FADING CHANNELS

Mohammad Javad Ahmadi
Ph.D. in Electrical and Electronics Engineering
Advisor: Tolga Mete Duman
January 2024

Random access techniques play a crucial role in machine-type communications (MTC), especially in the context of massive and sporadic device connectivity. Unlike traditional communication systems with scheduled access, random access allows devices to independently access the network without prior coordination. This flexibility is particularly beneficial for MTC scenarios where a large number of devices may transmit data sporadically. Unsourced random access (URA) is a form of grant-free random access in which devices remain entirely unidentified. As a result, there is no need for a codebook to store device identity preambles, whose dimension is squared to the number of connected users. This elimination of the codebook requirement empowers URA to efficiently accommodate an unbounded number of devices, reaching hundreds of millions of devices.

This thesis proposes three unsourced random access algorithms suitable for Gaussian and wireless fading channels. First, we introduce a URA algorithm for use over Gaussian multiple access channels. In the proposed solution, the users are randomly separated by assigning varying levels of transmit power to each of them. This introduces power diversity, enhancing the system performance. In the second part, we offer a solution for URA over Rayleigh block-fading channels with a receiver equipped with multiple antennas. We employ a slotted structure with multiple stages of orthogonal pilots; each randomly picked from a codebook. In the proposed signaling structure, each user encodes its message using a polar code and appends it to the selected pilot sequences to construct its transmitted signal. The receiver employs an iterative algorithm to detect messages transmitted by different users. This algorithm comprises several components, including pilot detection, channel estimation, soft data detection, single-user polar decoder, and successive interference cancellation. Additionally, we improve this scheme by incorporating an efficient strategy that separates users by random grouping. Our extensive analytical and simulation results demonstrate the effectiveness of

the proposed algorithm in terms of both energy efficiency and computational complexity. In the last part of the thesis, we study URA employing a passive reconfigurable intelligent surface, facilitating connections between the users and the base station when the direct link is blocked or significantly attenuated. We demonstrate through extensive simulations and analytical results that the proposed approach notably enhances system performance, particularly in channels with significant attenuation.

*Keywords:* Unsourced random access (URA), internet of things (IoT), grant-based access, grant-free access, uncoordinated access, coordinated access, orthogonal pilots, massive MIMO, pilot detection, power diversity, reconfigurable intelligent surface (RIS), RIS phase-shift design, Saleh-Valenzuela model, spreading sequence, massive MIMO, pilot detection, channel estimation, Log-likelihood ratio (LLR) generation, minimum mean squared error (MMSE), least squares (LS), maximum ratio combining (MRC), polar code.

# ÖZET

# GAUSS VE SÖNÜMLEMELİ KANALLAR İÇİN YENİ KAYNAKSIZ RASTGELE ERİŞİM ALGORİTMALARI

Mohammad Javad Ahmadi
Elektrik ve Elektronik Mühendisliği, Doktora
Tez Danışmanı: Tolga Mete Duman
Ocak 2024

Rastgele erişim teknikleri, özellikle büyük ve düzensiz cihaz bağlantısı bağlamında makine tipi iletişimlerde (MTC) önemli bir rol oynar. Zamanlanmış erişime sahip geleneksel iletişim sistemlerinin aksine, rastgele erişim, cihazların önceden koordinasyon olmaksızın bağımsız olarak ağa erişmelerine izin verir. Bu esneklik, özellikle birçok cihazın veriyi düzensiz olarak iletebileceği MTC senaryoları için faydalıdır. Kaynaksız rastgele erişim (URA), cihazların tamamen tanımlanmadığı bir hibesiz rastgele erişim biçimidir. Bu nedenle, cihaz kimlik girişlerini depolamak için boyutu kullanıcı sayısının karesiyle orantılı olan bir kod kitabına ihtiyaç duyulmaz. Bu kod kitabı gereksiniminin ortadan kaldırılması, URA şemasının etkili bir şekilde sayısı yüz milyonlara ulaşıp sınırsız sayılabilecek kadar cihazı kolayca barındırabilmesini sağlar.

Bu tez, Gauss ve kablosuz sönümleme kanalları için uygun üç kaynaksız rastgele erişim algoritması önermektedir. İlk olarak, Gauss çoklu erişim kanalları üzerinde kullanılmak üzere bir URA algoritması tanıtıyoruz. Önerilen çözümde, kullanıcılar, her birine değişen seviyelerde iletim gücü atanarak rastgele ayrılır. Bu, sistem performansını artıran güç çeşitliliği getirir. İkinci bölümde, çoklu antenlere sahip bir alıcı ile Rayleigh blok-sönümleme kanalları üzerinde URA için bir çözüm sunmaktayız. Çok aşamalı, her biri bir kod kitabından rastgele seçilmiş dikey pilotların kullanıldığı bir yuvalı yapıyı benimsemekteyiz. Önerilen sinyalleşme yapısında, her kullanıcı mesajı bir kutupsal kod kullanarak kodlanmakta ve bu iletim sinyalini oluşturmak için seçilen pilot dizilerine eklenmektedir. Alıcı, farklı kullanıcılardan iletilen mesajları algılamak için bir yinelemeli algoritma kullanır. Bu algoritma, pilot algılama, kanal kestirimi, yumuşak veri algılama, tek kullanıcılı kutupsal çözücü ve ardışık müdahale iptali gibi çeşitli bileşenleri içerir. Ayrıca, bu şemayı rastgele gruplandırma stratejisi ile kullanıcıları ayırmak için etkili bir strateji ile iyileştiriyoruz. Kapsamlı analitik ve simülasyon sonuçlarımız, önerilen algoritmanın enerji verimliliği ve hesaplama karmaşıklığı

açısından etkinliğini göstermektedir. Tezin son bölümünde, doğrudan bağlantı engellendiğinde veya önemli ölçüde zayıflatıldığında kullanıcılar ile baz istasyonu arasındaki bağlantıyı kolaylaştırmak için pasif bir yeniden yapılandırılabilir akıllı yüzey kullanıldığı bir URA şeması incelenmektedir. Simülasyonlar ve analitik bulgular ile önerilen yaklaşımın, özellikle önemli sönümleme yaşanan kanallarda, sistem performansını dikkat çekecek miktarda artırdığını gösteriyoruz.

*Anahtar sözcükler*: Kaynaksız rastgele erişim (URA), Nesnelerin İnterneti (IoT), hibeli erişim, hibesiz erişim, koordinatsız erişim, koordineli erişim, dik pilotlar, kapsamlı MIMO, pilot algılama, güç çeşitliliği, yeniden yapılandırılabilir akıllı yüzey (RIS), RIS faz kaydırma tasarımı, Saleh-Valenzuela modeli, yayılma dizisi, kanal kestirimi, Log-likelihood oranı (LLR) üretimi, Minimum Ortalama Kare Hatası (MMSE), En Küçük Kareler (LS), Maksimum Oran Birleştirme (MRC), polar kod.

# Acknowledgement

I want to extend my heartfelt appreciation to Prof. Tolga M. Duman for his valuable guidance and unwavering support. I have gained a wealth of knowledge from him, both in terms of scientific insights and personal growth, and I wish to convey my immense appreciation for this. I am profoundly grateful to my PhD committee (TIK) members, Prof. Sinan Gezici and Prof. Tolga Girici, for their insightful suggestions during our committee meetings. I would also like to express my gratitude to Prof. Ezhan Karaşan and Prof. Melda Yüksel for graciously agreeing to serve as jury members for my PhD thesis defense. Additionally, I want to extend my thanks to Dr. Mohammad Kazemi for engaging in fruitful discussions and collaborations on various research ideas.

I am grateful to acknowledge the support provided by the Scientific and Technical Research Council of Turkey (TÜBİTAK) under grants 119E589 and 221N366 for this work.

I extend my heartfelt gratitude to all the current and former members of the Bilkent Communication Theory and Application Research (CTAR) lab. Specifically, I want to express my thanks to my friends and colleagues who have shared memorable moments of my life: Mohammad Kazemi, Mert Ozates, Sadra Seyed-Masoumian, Muhammad Atif Ali, Busra Tegin, Reza Asvadi, Furkan Bagci, Uras Kargı, and Ruslan Morozov.

I would like to convey my eternal gratitude to my parents and brother for their unwavering love and support. Lastly, but certainly not least, I want to extend my deepest gratitude to my first and lifelong love, Faeze. Throughout the past decade, she has been much more than just a best friend; she has been a source of sacrifice, unwavering love, extraordinary companionship, and endless encouragement. Her presence has played a pivotal role in helping me navigate through this significant stage of my life.

# Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| AMP | Approximate message passing |
| BP | Belief propagation |
| BPSK | Binary shift keying |
| BS | Base station |
| CDMA | Code-division MAC |
| CCS | Coded/coupled compressive sensing |
| CRC | Cyclic redundancy check |
| CSI | Channel state information |
| CS | Compressed sensing |
| DOF | Degrees of freedom |
| FDMA | Frequency-division MAC |
| FEC | Forward error correction |
| GMAC | Gaussian multiple-access channel |
| IDMA | Interleave-division multiple access |
| i.i.d. | Identically distributed |
| IISD | Iterative inter-symbol decoder |
| IOT | Internet of Things |
| LDPC | Low-density parity-check |
| LLR | Log-likelihood ratio |
| LS | Least squares |

| | |
|---|---|
| MAC | Multiple access |
| MTC | Machine Type Communications |
| MIMO | Multiple-input multiple-output |
| ML | Maximum likelihood |
| MMSE | Minimum mean-square error |
| MSE | Mean square error |
| MS-MRA | Multi-stage set-up with multiple receive antennas |
| MS-MRA-WOPBE | MS-MRA without pilot bits encoding |
| MSUG-MRA | Multi-stage set-up with user grouping for multiple receive antennas |
| MRC | Maximum-ratio combining |
| NOMA | Non-orthogonal MAC |
| NP | Neyman-Pearson |
| OFDM | Orthogonal frequency-division multiplexing |
| OMP | Orthogonal matching pursuit |
| PMF | Probability mass function |
| PUPE | Per-user probability of error |
| QPSK | Quadrature phase shift keying |
| RS | Reed-Solomon |
| RIS | Reconfigurable intelligent surface |
| SDMA | Space-division MAC |
| SIC | Successive interference cancellation |
| SINR | Signal-to-interference-plus-noise ratio |
| TDMA | Time-division MAC |
| TIN | Treat interference as noise |
| URA | Unsourced random access |

# Chapter 1

# Introduction

Machine Type Communications (MTC) refers to the communication paradigm tailored for the interaction between devices without direct human intervention [1, 2]. It is a key aspect of the Internet of Things (IoT) paradigm, enabling various machines, sensors, and devices to exchange information [1–3]. MTC plays a crucial role in many applications such as smart cities, industrial automation, and healthcare [2, 4, 6]. MTC exhibits distinctions from human-type communication in several aspects. One notable difference is the ability of MTC to handle a large number of devices with sporadic activity for each device [5–8]. Specifically, MTC is anticipated to accommodate a density of up to 1 Million devices per square kilometer, and these devices typically operate with low computational and storage capabilities, low duty cycles, and small payloads [6–8].

Multiple access (MAC) techniques provide the opportunity for different users to share common communication media simultaneously, thus playing a fundamental task in wireless communication systems [9]. Centralized access (referred to as coordinated access) and random access (known also as initial access and uncoordinated access) are two main categories of MAC technology [4, 5, 7–12]. Centralized access refers to a communication protocol or technique where access to a shared communication channel is controlled and managed by a central authority, which allocates distinct communication resources to each user and allows

them to transfer data. On the contrary, in the random access, users send their messages simultaneously as soon as they have information without the need for explicit coordination with the central unit [4, 10–14].

Coordinated Multiple Access stands as the predominant multiple access mode in 5G and preceding technologies [8]. Several fundamental coordinated access techniques have been investigated from 1G to 4G, including time-division MAC (TDMA), frequency-division MAC (FDMA), orthogonal frequency-division multiplexing (OFDMA), code-division MAC (CDMA), and space-division MAC (SDMA). These techniques belong to the large category of orthogonal multiple access (OMA) [8]. In the context of 5G networks, the utilization of non-orthogonal multiple access (NOMA) improves the capacity region compared to the conventional techniques [15]. To manage the communication of available users in a system, the mentioned techniques require appropriate scheduling protocols, which are efficient only in the case of few active users [12, 17, 19]. However, in order to accommodate a large number of users in massive MTC (in 6G), the coordinated protocols incur excessively high access latency and prohibitive signaling overhead [4, 12, 14, 16–19].

In contrast to the conventional centralized access, where the base station (BS) waits for the preamble from devices to allocate resources to them, in random access, users transmit their data without any coordination. Removing the need for scheduling results in some benefits, such as reducing the latency and signaling overhead, which makes the random access scheme interesting for serving massive MTC systems with many users [4, 14, 16–19]. ALOHA and grant-free MAC are two common random access schemes [19, 20].

A widely used random access scheme is the ALOHA multiple access protocol, which enables users to transmit data without prior coordination. In the original description given in [21], Abramson explained that in an ALOHA network, a single broadcast channel is used by multiple communication devices, each of which independently sends packets without coordinating with other devices or following a specific schedule. Specifically, a device sends the entire packet at a randomly selected time, and then waits for an acknowledgment for a period of time. If

no acknowledgment is received within the timeout period, it is assumed that a collision has occurred with a packet sent by another device. In such cases, the packet is retransmitted after an additional random waiting time to avoid repeated collisions. This process is repeated until a successful transfer and confirmation occurs or until the process is terminated by the user console. Slotted ALOHA (SA), a variation of ALOHA, is presented as a method for transmitting packets in designated time slots, which doubles the maximum throughput compared to the pure ALOHA protocol [24]. However, there are still some challenges in SA. For instance, in this protocol, several active users may share the same slot to send their data packets, leading to packet collisions within that slot and discarding of data due to its irreplaceable nature [21–23]. Many studies have focused on improving the effectiveness of SA through various methods such as multiple packet transmission, interference cancellation, and error correction codes [25–28]. Also, further analyses of these techniques are provided in [29–32].

Grant-free random access is another type of random access method [6, 12]. In contrast to ALOHA, where users solely transmit data, the grant-free method involves each user transmitting both a preamble and the data sequence [20]. Preamble transmission aims to facilitate collision reduction at the physical layer. Hence, unlike ALOHA, which experiences a significant decline in performance during intense collisions, grant-free random access is more resilient to such collision scenarios [19, 33]. So, when there are massive number of users, grant-free random access is a more effective solution than ALOHA. Sourced and unsourced random access schemes are the main categories of grant-free random access [34]. In the first one, both the messages and the identities of users matter to the BS. Therefore, a distinct pilot/preamble is defined for each user, leading to the requirement of a codebook with dimensions proportional to the square of the potential user count. It is impractical specially in next-generation wireless networks with hundreds of millions of connected devices [34].

In the so-called *unsourced random access* (URA), which was introduced by Polyanskiy in [35], the focus is only on the messages transmitted, and the identity of the users is not important for the BS [36]. Removing user identity in the URA system enhances its efficiency compared to sourced random access, enabling it

to support a significantly larger number of users. In this paradigm, the BS is connected to millions of cheap devices, a small fraction of which are active at any given time. In this set-up, users share all resources without any one of them having priority over the others. The per-user probability of error (PUPE) is adopted as the performance criterion in URA system [36–39].

To support an even greater number of users in URA, various efficient technological solutions, such as multiple input multiple output (MIMO) systems, are also adopted [37, 38]. MIMO systems help URA by providing high spectral efficiencies, high energy efficiencies, high data rates, and spatial multiplexing gains. This is achieved by generating a substantial number of spatial degrees of freedom (DoF) [40–43].

Another promising technology that can furnish the URA system with high spectral efficiency and energy savings is the reconfigurable intelligent surface (RIS) [34, 39]. Specifically, a passive RIS equipped with many low-cost passive elements, which can intelligently tune the phase-shift of the incident electromagnetic waves, and reflect them in a desired direction without any amplification, can improve the efficiency of the network by enabling line-of-sight paths between the transmitters and the receivers in problematic environments with many blocking obstacles [34, 39, 44, 45]. URA schemes in the literature consider direct links between all the users and the BS; however, in certain environments, the direct link between some users and the BS may be blocked or significantly attenuated. Therefore, the use of RIS can improve user connectivity in URA by creating high-quality links between the BS and the users [34, 39].

## 1.1 Contributions of the Thesis

The objective of this thesis is to formulate novel and highly effective encoding and decoding blocks tailored to the requirements of unsourced random access. To this end, we develop an efficient URA coding scheme designed for use in GMAC. We also propose a URA scheme over Rayleigh block-fading channels with a receiver

equipped with multiple antennas. Furthermore, we introduce a URA scheme designed for scenarios where the direct communication link between users and the BS is obstructed. Our contributions within each of these three key areas are outlined as follows.

**URA with power diversify over GMAC:** In this line of work, we propose a random spreading approach with polar codes for unsourced multiple access, for which each user first encodes its message by a polar code, and then the coded bits are spread using a random spreading sequence. The proposed approach defines different groups, and employs different power levels for each group in such a way that the average power constraint is satisfied. We formulate and solve an optimization problem to determine both the number of groups and their respective power levels, along with the expected number of users within each group. Extensive simulations show that the proposed approach outperforms the existing methods, especially when the number of active users is large.

The results of this line of investigation have been published in [36].

**MIMO URA technique with multiple stages of orthogonal pilots:** We study the problem of URA over Rayleigh block-fading channels with a receiver equipped with multiple antennas. We propose a slotted structure with multiple stages of orthogonal pilots, each of which is randomly picked from a codebook. In the proposed signaling structure, each user encodes its message using a polar code and appends it to the selected pilot sequences to construct its transmitted signal. Accordingly, the transmitted signal is composed of multiple orthogonal pilot parts and a polar-coded part, which is sent through a randomly selected slot. The performance of the proposed scheme is further improved by randomly dividing users into different groups each having a unique interleaver-power pair. We also apply the idea of multiple stages of orthogonal pilots to the case of a single receive antenna. In all the set-ups, we use an iterative approach for decoding the transmitted messages along with a suitable successive interference cancellation technique. The use of orthogonal pilots and the slotted structure lead to improved accuracy and reduced computational complexity in the proposed set-ups, and make the implementation with short blocklengths more viable. Performance

of the proposed set-ups is illustrated via extensive simulation results which show that the proposed set-ups with multiple antennas perform better than the existing MIMO URA solutions for both short and large blocklengths, and that the proposed single-antenna set-ups are superior to the existing single-antenna URA schemes.

These research outcomes have been published in [37, 38].

**RIS-aided URA scheme:** This study considers a URA set-up equipped with a passive RIS, where a massive number of unidentified users (of which only a small fraction are active at a given time) share the same communication resources. We propose a slotted transmission scheme that operates in two phases. In the first phase, called the RIS configuration phase, the BS detects the active pilots and estimates their respective CSI. Then, using the estimated CSI, the BS suitably selects the phase shifts of the RIS elements. In the second phase, called the data phase, transmitted messages of active users are decoded. The proposed channel estimator offers the capability to estimate the channel coefficients of the users whose pilots interfere with each other without prior access to the list of selected pilots or the number of active users. In the design of RIS elements, we introduce an algorithm that relies on the semidefinite relaxation technique, as well as a more efficient alternative based on eigenvalue decomposition. The latter offers comparable performance while maintaining the high accuracy of the former method. This study explores two scenarios: one in which the direct link between users and the BS is entirely blocked, and the other in which direct links exist between them. We demonstrate that improving the connectivity between the BS and users through RIS enhances the performance of the URA system, particularly in situations where the direct link is obstructed or experiences substantial signal attenuation. The effectiveness of the proposed algorithms is confirmed through extensive numerical evaluations and simulations results.

Part of our results on this topic have been presented at IEEE GLOBECOM 2023 [39].

## 1.2 Thesis Outline

The rest of the thesis is organized as follows. Chapter 2 provides an overview of unsourced random access and presents a literature survey, considering both GMAC and fading channels. Chapter 3 introduces an efficient coding scheme that utilizes power diversity to enhance the performance of the URA system on a GMAC channel. In Chapter 4, we turn our attention to the URA over MIMO Rayleigh fading channel by presenting a pilot-based scheme with multiple stages of orthogonal pilots, improving it by randomly grouping the users, and providing a theoretical analysis on its performance. In Chapter 5, we introduce a RIS-assisted URA scheme designed to enhance system performance in scenarios where the direct links between the base station and users are significantly attenuated. Finally, we summarize our results, and provide conclusions and directions for future research in Chapter 6.

## 1.3 Notation

Lower-case and upper-case boldface letters are used to denote a vector and a matrix, respectively; We denote the sets of real and imaginary numbers by $\mathbb{R}$ and $\mathbb{C}$, respectively; $\text{diag}(\mathbf{t})$ and $\mathbf{I}_N$ represent a diagonal matrix with elements of vector $\mathbf{t}$ in its diagonal, and an $N \times N$ identity matrix, respectively; $\text{Re}(\cdot)$, $\text{Im}(\cdot)$, and $\text{trace}(\cdot)$ denote the real and imaginary parts, and trace of a matrix, respectively; $[\mathbf{T}]_{(i,j)}$ refers to the element in the $i$th row and $j$th column of $\mathbf{T}$; $[\mathbf{T}]_{(l,:)}$ and $[\mathbf{T}]_{(:,l)}$ represent the $l$th row and the $l$th column of $\mathbf{T}$, respectively; $[\mathbf{t}]_i$ represents the $i$th element of the vector $\mathbf{t}$; $\text{vec}(\cdot)$ is the vectorization operator; $\otimes$ denotes the Kronecker product; $\mathcal{CN}(\mathbf{0}, \mathbf{B})$ denotes the zero-mean circularly symmetric complex Gaussian random variable with covariance matrix $\mathbf{B}$; $U(c, d)$ is the continuous uniform distribution on $[c, d]$; the transpose and Hermitian of the matrix $\mathbf{T}$ are denoted by $\mathbf{T}^T$ and $\mathbf{T}^H$, respectively; $|.|$ denotes the cardinality of a set, $\mathbf{I}_M$ and $\mathbf{1}_s$ denote the identity matrix and $1 \times s$ all-ones vector, respectively; we use $[a_1 : a_2]$ to denote $\{i \in \mathbb{Z} : a_1 \leq i \leq a_2\}$; $\delta_{i,j}$ is the Kronecker delta; and, the set $\mathcal{F}(N) = \{-1, -1 + \Delta_N, -1 + 2\Delta_N, ..., -1 + (N-1)\Delta_N\}$ with $\Delta_N = 2/N$.

# Chapter 2

# Preliminaries and Literature Review

In this chapter, we provide an overview of contemporary literature related to URA, with a focus on GMAC, fading MAC, and MIMO fading MAC. More specifically, we study the works that put forth coding schemes and those that offer fundamental limits for the URA setup. Considering a URA model with $M$ receive antennas, in which $K_a$ out of $K_T$ users are active at a given frame, the received signal matrix in the absence of synchronization errors is written as [36–38]

$$\mathbf{Y} = \sum_{i=1}^{K_a} \mathbf{h}_i \mathbf{x}_i(\mathbf{u}_i) + \mathbf{Z} \in \mathbb{C}^{M \times n}, \tag{2.1}$$

with

$$\begin{cases} \mathbf{h}_i = 1, \quad M = 1 & \text{GMAC} \\ \mathbf{h}_i \quad \sim \mathcal{CN}(0,1), \quad M = 1 & \text{Fading channel} \\ \mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M), \quad M > 1 & \text{MIMO fading channels} \end{cases} \tag{2.2}$$

where $\mathbf{x}_i(\mathbf{u}_i)$ is the length $n$ signal corresponding to the message bit sequence $\mathbf{u}_i \in \{0,1\}^B$ of user $i$, $B$ denotes the number of information bits, and elements of the additive Gaussian noise $\mathbf{Z} \in \mathbb{C}^{M \times n}$ are drawn from $\mathcal{CN}(0, \sigma_z^2)$. Each user selects its message index uniformly from the set $\{1, 2, ..., 2^B\}$. The objective

in URA is to determine $\mathbf{u}_i$ $\forall i = 1, 2, ..., K_a$ by analyzing the received signal $\mathbf{Y}$ [35–38].

## 2.1 Unsourced Random Access: Challenges and Potential Solutions

To address specific constraints within the URA paradigm, certain challenges need to be overcome, e.g., 1) heavy interference: the grant-free nature of communication leads to the received signal being a superposition of signals from multiple users, resulting in considerable interference for each individual user; 2) collisions: since users are unidentified, they have to randomly share communication resources, which can result in heavy collisions; 3) developing a multi-user coding scheme: decoding in massive random access techniques is a complex task, unlike the single-user scenarios where various robust channel codes are available, 4) Computational complexity: the task of separating and decoding a massive number of users involves substantial computational complexity. Consequently, presenting decoding algorithms with low complexity is an important task in URA. When designing a URA system, it is crucial to seek effective solutions to address these challenges. In the following, we will delve into these challenges along with possible solution approaches.

- **Development of multi-user coding schemes:** For decoding different users' messages with the received signal in (2.1), the most immediate solution is that active users map their bit sequences to the rows of a $2^B \times n$ codebook to select their transmitted signals. The receiver can separate the signals of different users and detect each one using detectors such as maximum likelihood (ML) [35, 52]. Nevertheless, it is crucial to note that this coding scheme is only practical for lower values of $B$ because its computational complexity experiences exponential growth as $B$ increases.

  In the context of coded compressed sensing and tensor-based techniques (which will be discussed in the following sections), a more computationally

efficient approach is adopted [34, 53–59]. These methods involve dividing each user's bit sequence into distinct segments, allowing each segment to be mapped to a codebook with reduced dimensions. This reduction in dimensionality helps lower the computational load on the decoder, albeit at the cost of reduced accuracy.

In another category of coding schemes, some techniques are employed to separate the received signals of different users. Then, each user can encode its whole message sequence (without partitioning it) with a strong single-user channel code such as polar, low-density parity-check (LDPC), forward error correction (FEC), etc [36–39, 60–79]. These algorithms are considered efficient because they manage to strike a balance, avoiding both the high computational complexity associated with ML-based coding schemes and the low accuracy observed in schemes that segment bit sequences.

- **Dealing with heavy interference:** Interference stands out as a primary challenge in multiple access channels. Failure to incorporate strategies for reducing or canceling it can lead to a significant decline in performance. To demonstrate the detrimental impact of interference, we study the performance of the treat interference as noise (TIN) scheme, where single-user decoding is carried out by treating interference from other users as noise, without employing any techniques to mitigate it. Approximating $\|\mathbf{h}_i\|^2 \approx \|\mathbf{h}_j\|^2 \ \forall i, j$ in (2.1), the signal-to-interference-plus-noise ratio (SINR) of a user in the TIN strategy can be written as [36]

$$\alpha_i \approx \frac{P}{\sigma_z^2 + (K_a - 1)P}, \tag{2.3}$$

where $P$ is the transmitted power of each user. For $K_a$ being large enough, the SINR can be written as

$$\alpha_i \approx \frac{1}{K_a - 1}. \tag{2.4}$$

This equation illustrates that when the number of users surpasses a specific threshold, the SINR exhibits a linear decrease concerning $K_a$, and boosting the transmitted power does not result in an enhanced SINR. Consequently, a significant rise in $K_a$ (equivalent to an increase in interference)

adversely impacts the system's performance. Furthermore, one can observe this phenomenon by consulting [80], where the TIN scheme is compared with TIN-SIC as a strategy designed to mitigate interference. The adverse effects of neglecting interference reduction in the TIN scheme are evident from the result of this study.

There are some techniques in the literature to mitigate the interference:

- Slotting: Dividing the time frame into multiple slots serves to diminish interference by reducing the number of users in each slot [37–39,80,81]. Additionally, slotting results in a shorter duration for the transmitted signal and a reduction in the number of users per slot. Consequently, the overall computational complexity of the URA system is decreased because an increase in these two parameters superlinearly increases the computational workload of the multi-user decoder [38, 77].
- Sending some preambles/pilots: In some URA schemes, each user randomly selects a preamble/pilot, and transmits it along with its main signal [36–39,62,63,72,73,77]. Upon detecting these transmitted preambles/pilots at the receiver, there is an opportunity to signify the influence of a desired user while mitigating the impact of other interfering ones. For instance, in certain MIMO fading URA schemes, each user selects a pilot and appends it to their codeword before transmission. At the receiver's end, the channel coefficients for different users can be estimated by using the detected pilots. With these estimated channel vectors for various users, interference can be significantly reduced by employing estimators like the minimum mean square error (MMSE) [37–39, 72, 73, 77].
- SIC: Successive interference cancellation (SIC) is a useful technique for interference mitigation in wireless multi-user scenarios. The idea behind it is to re-encode the successfully detected messages and remove them to decrease the interference [36–39, 83, 84]. It is important to note that in fading channels with unknown CSI, an additional step is necessary. This step involves estimating the corresponding channel coefficients of a detected message before proceeding with the subtraction

process [37–39].

  - Power diversity: An alternative approach for user separation involves randomly distributing users into different groups with varying power levels and adopting an appropriate technique to ensure that the decoding process for each group is independent of the others [36, 38]. This approach serves to reduce interference within each group. As previously discussed in the context of slotting, dividing users into distinct groups also leads to a decrease in computational complexity of the URA system.

- **Addressing computational complexity:** To increase the number of active users supported by a URA system, larger values of parameters, such as frame length and the number of antenna elements in MIMO systems, are necessary [35, 56, 88]. Additionally, the computational complexity of the system escalates with these parameters and the number of active users, resulting in a superlinear dependence of computational load on the active user count [37, 77]. Consequently, a URA system with a substantial number of active users is anticipated to exhibit high computational complexity. To mitigate this complexity, various strategies are employed, such as slotting the frame length and randomly dividing users into different independent groups [36–38].

- **Reducing collisions:** In multi-user systems, a collision occurs when two or more users access the same communication resource simultaneously. In a URA system, a certain degree of collision, such as multiple users sharing the same time slot or bandwidth, may be resolved. Nevertheless, there are cases where the system encounters failure in the event of a collision. For instance, in schemes where users choose from a preamble/pilot codebook [36, 39, 64, 72], if multiple users happen to share the same preamble/pilot, the system struggles to decode them. Therefore, it is crucial to perform a preliminary analysis to keep the collision probability lower than a desired threshold [38] or suggest techniques for its mitigation [37, 38].

| Paper | Contribution and techniques |
|---|---|
| Polyanskiy et al., 2019 [35] | Achievability and converse bounds + comparing with CDMA, TDMA, and FDMA. |
| Glebov et al., 2023 [52] | Achievability bound for binary and Gaussian codebooks showing comparable performance to [35]. |
| Vem et al., 2019 [60, 61] | A novel coding scheme that includes LDPC codes, slotting, permutation, preamble transmission, and SIC. |
| Pradhan et al., 2019 [62, 63] | A novel coding scheme that includes LDPC code, permutation, preamble transmission, and SIC. |
| Marshako et al., 2019 [67] | A novel coding scheme that includes polar code, slotting, and SIC. |
| Pradhan et al., 2020 [64] | A novel coding scheme that includes polar code, preamble transmission, random spreading, and SIC. |
| Pradhan et al., 2021 [65] | A novel coding scheme that includes LDPC code, preamble transmission, random spreading, and SIC. |
| Han et al., 2021 [66] | A novel coding scheme that includes FEC, preamble transmission, sparse spreading, and SIC. |

Table 2.1: Summary of available coding schemes in URA over GMAC.

## 2.2 URA over GMAC

In this section, we review some basic results on URA over GMAC. In [35], Polyanskiy established an achievability bound for unsourced random access, and compared it with practical approaches including orthogonalization (FDMA and TDMA), ALOHA, and CDMA, illustrating the superiority of URA over conventional methods. Additionally, [35] delves into the asymptotic coding challenges for a $K$-user GMAC, where $K$ scales with the blocklength and each user's payload remains constant. While the achievability bound discussed in [35] is only for Gaussian codebooks, the authors of [52] derive bounds for both Gaussian and binary codebooks, and demonstrate that such codebooks exhibit comparable performance to that given by Polyanskiy. Hence, they illustrate that Polyanskiy's achievability bound is not limited solely to Gaussian codebooks; instead, it holds for a broader range of signaling structures. Here, the binary codebook consists of entries drawn from {+1, -1}, while the Gaussian codebook involves entries drawn from $\mathcal{CN}(0, 1)$.

Several practical coding schemes have also been developed for URA over a GMAC [60–67]. Despite having significantly lower computational complexity compared to the coding scheme used to establish the achievability bound in [35], these coding schemes exhibit comparable performance in certain scenarios. The proposed solutions leverage a variety of methods to separate and then decode the signals of distinct users. For instance, in some of them, each user divides its bit sequence into two components: a preamble and a data part [60–66]. The first part is mapped to a codebook to select a unique interleaving pattern, preamble, or spreading sequence, which separates different users by giving identification to them. The data part is encoded by a powerful channel code, e.g., a polar code or an LDPC code. Another effective approach is slotting the frame [67]: as signals from other users can act as sources of interference for a particular user, randomly dividing users into distinct slots can reduce the number of users within each slot, thereby having lower interference level in each slot. SIC is another efficient technique for reducing the interference, which mitigates interference by eliminating successfully decoded codewords [60–67].

Table 2.1 provides a comprehensive summary of the techniques employed in various schemes over the GMAC channel. Additionally, Figure 2.1 examines their performance with respect to different number of active users. The results demonstrate that when the number of active users is low, some of low-complexity coding schemes achieve a performance similar to the achievability bound of Polyanskiy in [35]. Conversely, with an increasing number of active users, the performance gap becomes more pronounced. The improved performance observed in schemes employing spreading sequences in [64–66] can be attributed to the spreading of symbols from different users by distinct sequences, hence resulting in reduced effective interference. In these schemes, the received signal associated with each symbol is a mixture of $K_a$ vectors, rather than $K_a$ scalar values. This introduces an additional degree of freedom that simplifies the procedure of estimating symbols for individual users. This, in turn, provides us with a way of generating LLR values with minimal interference from other users. In contrast, the schemes in [60–63, 67] generate LLRs without making any effort to estimate the symbols beforehand.

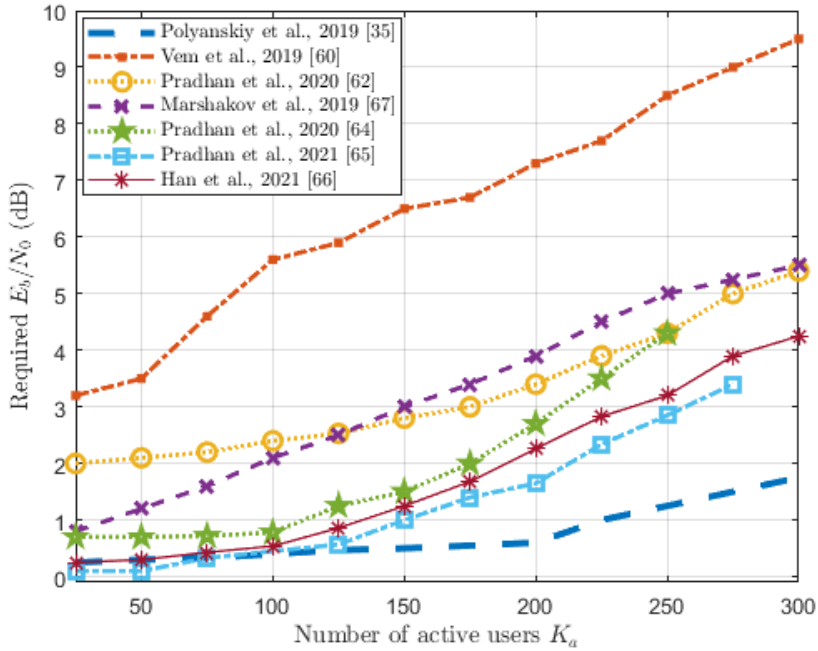


Figure 2.1: The required $E_b/N_0$ to achieve the target PUPE of 0.05 as a function of the number of active users for $n = 30000$, $B = 100$, and different coding schemes over GMAC channel.

## 2.3 URA over Fading Channels

Since GMAC is not a very realistic channel model for wireless communications, a number of studies explore URA over Rayleigh fading channels [54, 63, 68–71, 80–82, 85–87]. In the references [80–82, 85, 86], the authors derive some performance bounds for the URA set-up over a Rayleigh fading channel. In [81, 82], achievability and converse bounds are proposed for the cases of known and unknown CSI in an asymptotic setting, where the number of users grows linearly with the blocklength. Authors in [85, 86] provide both asymptotic and non-asymptotic achievability and converse bounds for the case of unknown CSI. In [80], authors derive an achievability bound for URA scheme over Rayleigh fading channel by considering slotted structure and perfect SIC in their analysis. As clear from Figure 2.6, their proposed non-asymptotic achievability bound works better than

the ones in [81, 82, 85, 86] (i.e., its energy efficiency performance is closer to the converse bound), however, the assumption of perfect SIC is not realistic.

The references [53, 54, 63, 68–71, 80, 85–87] introduce practical coding schemes tailored for the URA scheme on fading channels. Within this collection of works, some assume that the received signals from all users are synchronized [53, 54, 63, 80, 85, 86], while others take into account asynchronous transmission [68–71, 87]. To estimate the delay of asynchronous signals, [68–70,87] use OFDM transmission to convert the time-delay to the phase shift in the frequency domain, while the authors in [71] consider time-domain transmission, where the delay tap of a signal is detected by finding the delay at which the compressed sensing (CS) preamble has a peak. A summary of URA studies over fading channels involving a single receive antenna is given in Table 2.2, and the performances of different coding schemes are compared in Figure 2.6.

Focusing on the coding schemes in the context of Rayleigh fading channels, we have the following classification:

- **Treating interference as noise (TIN)**: Among all the URA algorithms, this particular scheme is perhaps the simplest but it exhibits the poorest performance. It operates under the assumption of the presence of just a single user and treats other interfering users as noise. Consequently, this approach results in an exceptionally low average signal-to-noise ratio (SNR) due to the fact that it encounters interference-plus-noise as the system's overall noise level. Its performance compared to other URA schemes over the Rayleigh fading channel is clear from Figure 2.6.
- **Coded/coupled compressed sensing (CCS)** [53, 54, 71]:
  - Encoder: the encoding process is performed in two phases (Figure 2.2 top): 1) in the tree encoder, $B$ bits of information are divided into $S$ parts, and all of these segments, except the initial one, are appended by additional redundant parity-check bits, 2) in the CS encoder, each of $S$ sub-message is mapped to a CS codebook to select a sequence for

| Paper | Contribution | Description | Synchron. vs. Asynch. |
|---|---|---|---|
| Kowshik et al., 2019&2021 [81, 82] | Bounds | Achievable and converse bounds for known and unknown CSI in an asymptotic setting + comparing with CDMA, TDMA, and FDMA. | Synchron. |
| Andreev et al., 2020 [80] | Bounds & coding scheme: polar code with SIC | Non-asymptotic achievability and converse bounds for unknown CSI + a new coding scheme containing slotting, polar code, and SIC. | Synchron. |
| Kowshik et al., 2019&2020 [85, 86] | Bounds& LDPC code with joint decoding | Achievable and converse bounds for known and unknown CSI in asymptotic and non-asymptotic manners + a new coding scheme containing LDPC code, slotting, and iterative joint decoding algorithm. | Synchron. |
| Amalladinne et al., 2020 [53] | Coding scheme: CCS-based | A novel coding scheme consists of pilot transmission and detection and tree code. | Synchron. |
| Andreev et al., 2022 [54] | Coding scheme: CCS-based | Improving CCS in [53] by replacing the conventional outer tree code with 1) a code capable of correcting $t$ errors, and 2) a Reed–Solomon code. | Synchron. |
| Pradhan et al., 2022 [63] | Coding scheme: Sparse IDMA | A novel coding scheme that includes LDPC code, signal repetition, permutation, preamble transmission, and SIC. | Synchron. |
| Andreev et al., 2019 [68, 69] | Coding scheme: LDPC code with SIC | A new coding scheme containing LDPC code, OFDM, slotting, and SIC. | Asynch. |
| Chen et al., 2017 [70] | Coding scheme: OFDM with SIC | A new coding scheme containing OFDM and SIC. | Asynch. |
| Ozates et al., 2023 [87] | Coding scheme: OFDM with SIC | A new coding scheme in the frequency-selective channel, containing OFDM, pilot transmission, slotting, polar code, and SIC. | Asynch. |
| Amalladinne et al., 2019 [71] | Coding scheme: CCS-based | They extend CCS in [53] to accommodate asynchronous scenarios. | Asynch. |

Table 2.2: Summary of available coding schemes in URA over fading channels.

transmission. Hence, the received signal is comprised of $S$ slots, with each slot being the superposition of sequences from $K_a$ users.

- Decoder: the decoder also functions in two stages. In the initial stage, $K_a$ sequences associated with the $K_a$ active users are identified within each slot. Subsequently, with the assistance of the unique parity bits, the sub-messages corresponding to the detected sequences are combined to create a message of length $B$ for each user (the second stage is called outer decoder or tree decoder which is shown in Figure 2.2 bottom).

- Variations: The explanations above are for the CCS algorithm in [53], which is extended in other studies [54, 71]. Although [53] is designed for the GMAC case, it can be applied to URA over fading channels as well.

  In [71], the authors utilize CCS for the asynchronous URA model. More specifically, during the encoding stage, they append $T$ additional zeros to the end of each CS sequence, where $T$ represents the maximum possible delay of the system. At the receiving end, the decoder takes into account all potential delays, ranging from 1-symbol to a maximum of $T$-symbol delays, when performing CS decoding.

  Furthermore, in [54], authors modify the tree decoder of the CCS scheme in [53] by replacing it by two codes capable of correcting $t$ errors: the first one modifies the conventional tree code by giving it ability to recover $t$ errors (called $t$-tree), and in the second one, Reed-Solomon (RS) code is employed. As shown in Figure 2.6, $t$-tree scheme improves the performance of the CCS in the low $K_a$ regime. It is shown that, the use of RS codes improves both schemes significantly, especially in the case of small number of active users. Also, it can be seen in this figure that different variants of CCS exhibit lower energy efficiency compared to other schemes, except for TIN. The reason is that they employ a sub-optimal coding scheme, while others use strong channel codes such as LDPC or polar codes.

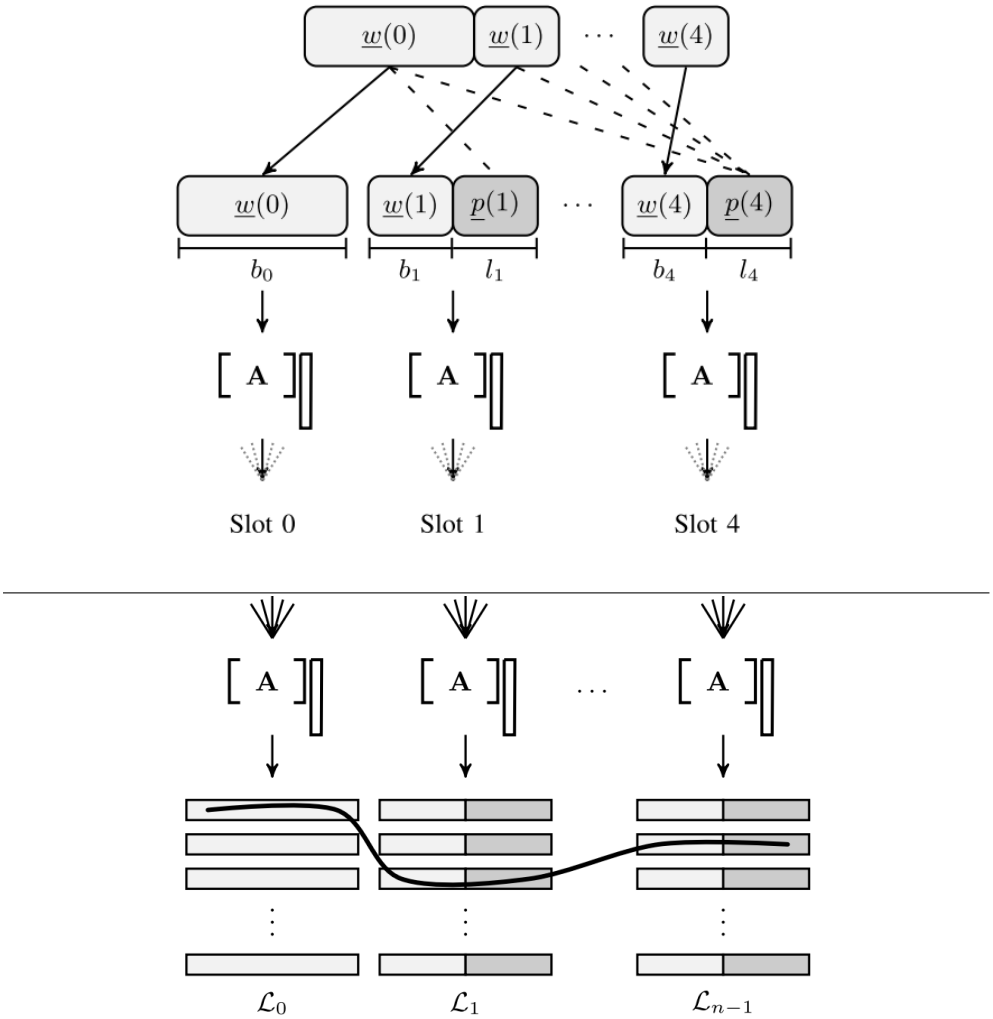


Figure 2.2: Encoding (top) and decoding (bottom) structures of a CCS scheme with 4 slots [53]. As depicted in the top figure, each user's signal is divided into 4 parts, and each part is appended to a parity-check bit. The result is mapped to a codebook to construct the pilot of the user in a slot. Also, at the receiver side, sub-messages of different users are detected at each slot, and corresponding ones in different slots are connected to form the complete message sequence of a user.

- **LDPC with joint decoding** [85, 86]:
  - Encoder: these algorithms partition the time frame into distinct slots. Each user transmits its LDPC codeword through a randomly chosen slot. As previously mentioned, slot allocation offers two significant

benefits: it reduces the computational load on the decoder by reducing the length of the LDPC signal and mitigates per-slot interference by distributing users across various slots.

– Decoder: to perform decoding, they utilize an iterative belief propagation (BP) decoder, as illustrated in Figure 2.3. The graph incorporates four distinct node types: variable nodes (in red), and check nodes (in blue), which constitute the traditional LDPC decoder; functional nodes (in green) corresponding to the symbols of the received signal; and a fourth type (in magenta) corresponding to the fading coefficients. This graph allows them to simultaneously update the estimated values of both the channel coefficients and the symbols for each individual user. The diminished performance observed in [85, 86] compared to the ones in [63, 68, 69, 80] (as illustrated in Figure 2.6), can be attributed to the absence of the highly efficient SIC block.

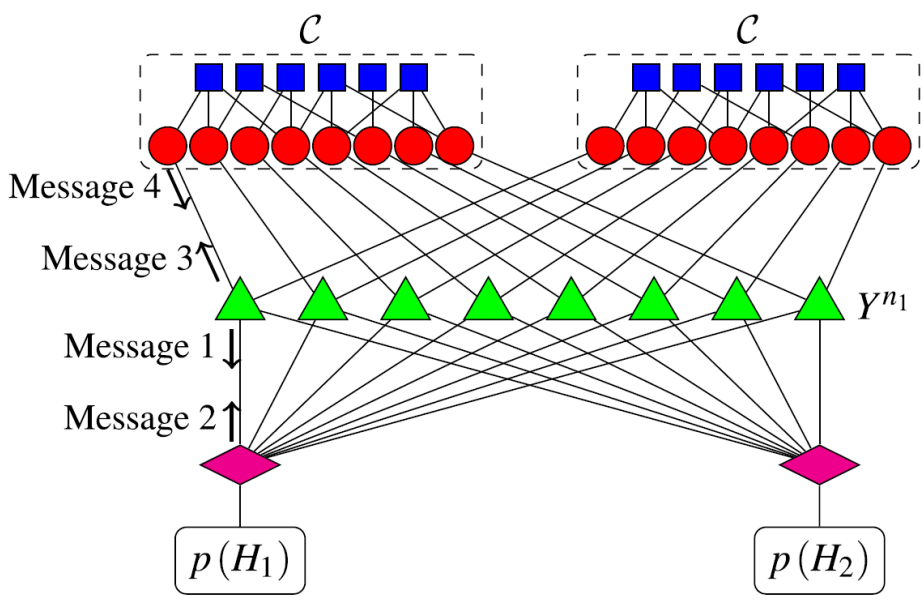


Figure 2.3: Iterative belief BP decoder for two users [85].

- **Sparse IDMA** [63]: Below, the encoding and decoding structures of sparse interleave-division multiple access (IDMA) are presented:

– Encoding: As illustrated in Figure 2.4, every user partitions its message bit sequence into preamble and data segments. The data portion of the message is subjected to LDPC encoding, repetition, and zero-padding to fill the entire frame length requirement, experiences permutation according to a predefined pattern, and ultimately gets attached to a preamble sequence before transmission. The preamble part of the message is used for selecting the preamble, the permutation pattern, and the number of repetitions.

– Decoding: due to synchronous transmission, the signal received at the receiver's end is a composite of signals from $K_a$ users. To decode the transmitted bits of each user, an initial step involves detecting the active preambles within the initial part (preamble part) of the received signal. This detection is accomplished using LASSO as a CS solver. Subsequently, by utilizing the identified preambles, the repetition count and permutation pattern are determined. These, in turn, determine the location of each symbol. Ultimately, the data bits are detected by employing a message-passing algorithm tailored for LDPC codes. As Figure 2.6 clearly illustrates, the approaches presented in [68, 69, 80] clearly surpass sparse IDMA. This performance difference arises from the necessity of allocating extra energy for transmitting preambles in sparse IDMA. In contrast, the former schemes distribute users randomly across different slots, obviating the need for a preamble to specify their locations.

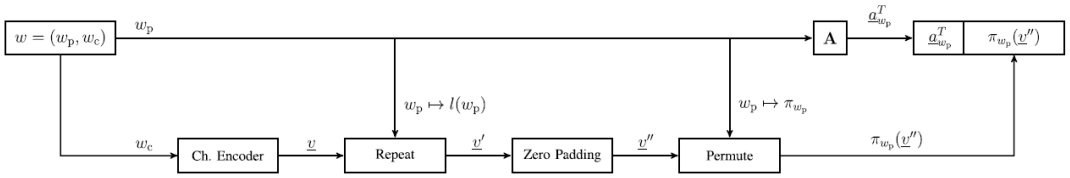


Figure 2.4: Encoding process of sparse IDMA [63].

- **LDPC coding with SIC** [68, 69]: This coding design builds upon the schemes presented in [85, 86], with the key distinction of incorporating OFDM techniques to manage asynchronous scenarios and employing SIC

instead of joint decoding. Performance of synchronous LDPC codes with SIC surpasses that of LDPC with joint decoding [85, 86], as clearly illustrated in Figure 2.6. This superior performance can be attributed to the enhanced capabilities of SIC compared to joint decoding.

- **Polar coding with SIC** [80]: Below, we will delve into the encoding and decoding procedures of this scheme, which stands out as the state-of-the-art among URA schemes operating over fading channels. The frame in this scheme is divided into $S$ slots. Each user randomly selects a slot to transmit its polar coded and modulated signal. For decoding, the scheme considers TIN strategy followed by an SIC: since different users in a slot are prone to different fading coefficients, some of them with larger coefficients are more likely to be decoded by TIN. After decoding the strongest user's signal while treating others as noise, its channel coefficient is estimated by OMP. Given the estimated channel coefficient and the decoded signal of a user, its contribution can be removed from the received signal using SIC. This strategy continues until no signal is successfully detected during an iteration. It is important to emphasize that the determination of whether a decoded message is successful or not relies on the use of a cyclic redundancy check (CRC) message sequence. The encoding and decoding process of the polar coding with SIC scheme in a slot with $K$ users is demonstrated in Figure 2.5.

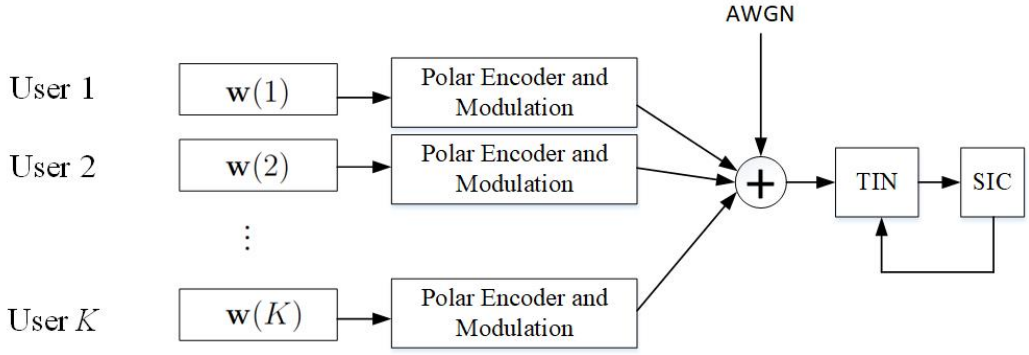


Figure 2.5: Encoding and decoding process of polar code with SIC in [80] in a slot with $K$ users.

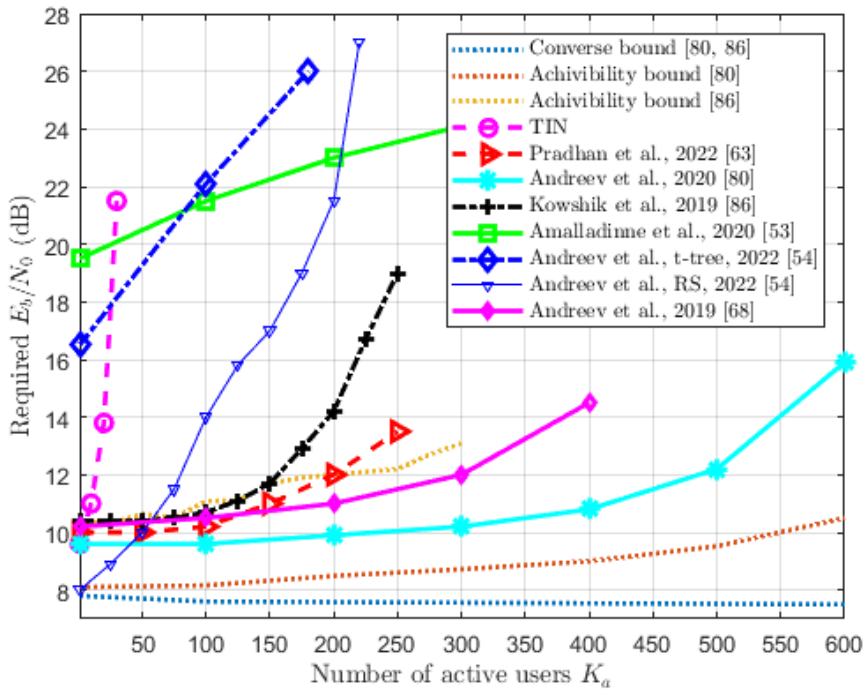


Figure 2.6: The required $E_b/N_0$ to achieve the target PUPE of 0.1 as a function of the number of active users for $n = 30000$, $B = 100$, and different coding schemes over Rayleigh fading channel.

## 2.4 URA over MIMO Fading Channels

In [88], achievability and converse bounds are proposed for the URA scheme over MIMO Rayleigh fading channels. Furthermore, several studies have investigated practical coding schemes for MIMO setting, which can be categorized into four groups: CCS-based, tensor-based, pilot-based, and Bayesian approaches [34, 56–59, 72–79].

As the computational complexity of mapping bit sequences of each user to a codebook increases exponentially with the parameter $B$, CCS-based algorithms are employed in which the bit sequence of each user is divided into $S$ sub-messages [56, 57]. Subsequently, these sub-messages are mapped to pilot codebooks with

lower dimensions (see Figure 2.2 top as an example with $S = 4$) to obtain $S$ different pilots. Then, pilots are transmitted through $S$ different slots. At the receiving end, $K_a$ sub-messages are detected within each slot, and they are then combined (see Figure 2.2 bottom) to form the complete length-$B$ bit sequence for each user. To establish connections between these sub-messages, the authors in [56] incorporate parity-check bits into each sub-message, while [57] relies on the correlation between channel coefficients of each user in different slots (assuming a quasi-fading channel model). Eliminating the necessity for additional parity bits enhances the performance of [57] in comparison to [56].

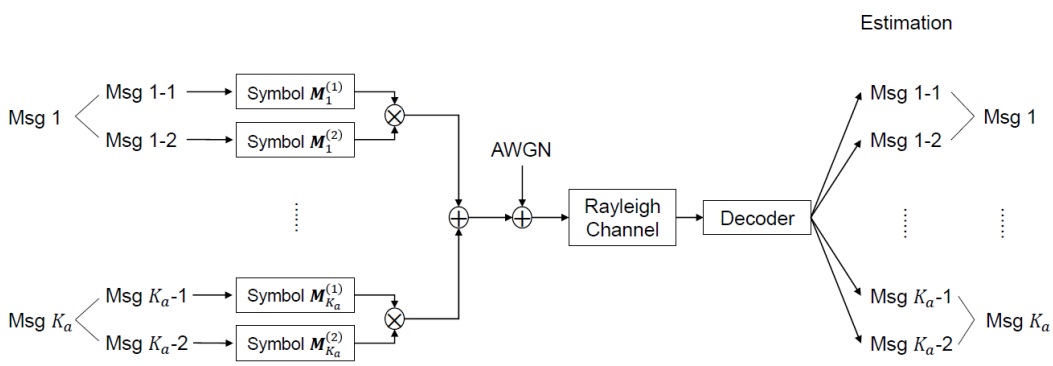


Figure 2.7: Encoding and decoding structures of tensor-based schemes, where each user divides its message sequence into $d$ parts (here $d = 2$), then applies Kronecker product on them [59].

In tensor-based URA schemes, each user sends a rank-1 tensor of order $d$. Therefore, the received signal is a rank-$K_a$ tensor of order $d + 1$ (considering the channel coefficient vector as an extra dimension of the tensor) summed with an additive noise [34, 58, 59]. For decoding, it is enough to only perform tensor decomposition on the received signal. In particular, by tensor decomposition, the rank-$K_a$ tensor of order $d+1$ is decomposed into $K_a$ rank-1 tensors of order $d+1$, which is equivalent to separating the signals of $K_a$ users, where the transmitted tensors can be identified by the first $d$ dimensions, and the $(d + 1)$th dimension gives the channel vector of the user. The schemes in [58, 59] address the challenge of decoding in URA with MIMO Rayleigh fading channels, while the authors in [34] attempt to employ a passive RIS to decode users in a scenario where there

are obstructed channels between the unsourced users and the BS. The overall coding structure of tensor-based methods is depicted in Figure 2.7.

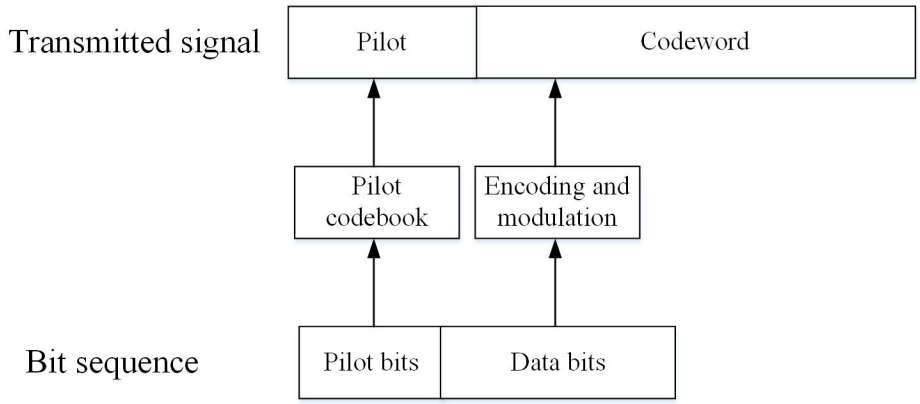


Figure 2.8: Encoding structure of pilot-based schemes.

Pilot-based techniques constitute another category of algorithms in the context of MIMO URA systems [72–77]. The idea behind them is to divide the bit sequences into pilot and data parts. The transmitted signal of each user also consists of two parts (see Figure 2.8 for details): the pilot part which is obtained by mapping the pilot bits to a codebook, and the data part which is the encoded (by polar or LDPC codes). For decoding, as shown in Figure 2.9, these schemes iteratively employ the following steps: 1) at the initial stage, they detect the active pilots using different algorithms such as orthogonal matching pursuit (OMP), approximate message passing (AMP), energy detector, etc., 2) the channel coefficient vectors corresponding to each detected pilot is estimated by algorithms like MMSE, 3) using the estimated channel coefficients, they obtain a soft estimation of the transmitted codewords, which are used for generating log-likelihood ratio (LLR) values, 4) after feeding the LLR to the decoder, the successfully decoded codewords are removed from the received signal for decreasing the interference level at next iterations (this step is not employed by all the pilot-based schemes). As shown in Figure 2.10, due to adopting strong channel coding algorithms such as LDPC and polar codes, pilot-based schemes usually outperform the tensor-based and CCS-based schemes for which sub-optimal coding schemes are employed.

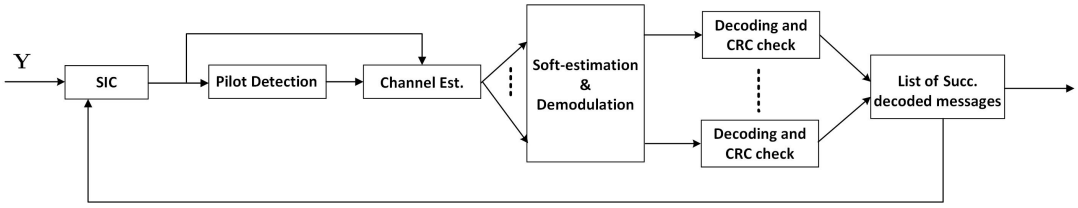


Figure 2.9: Decoding structure of pilot-based schemes.

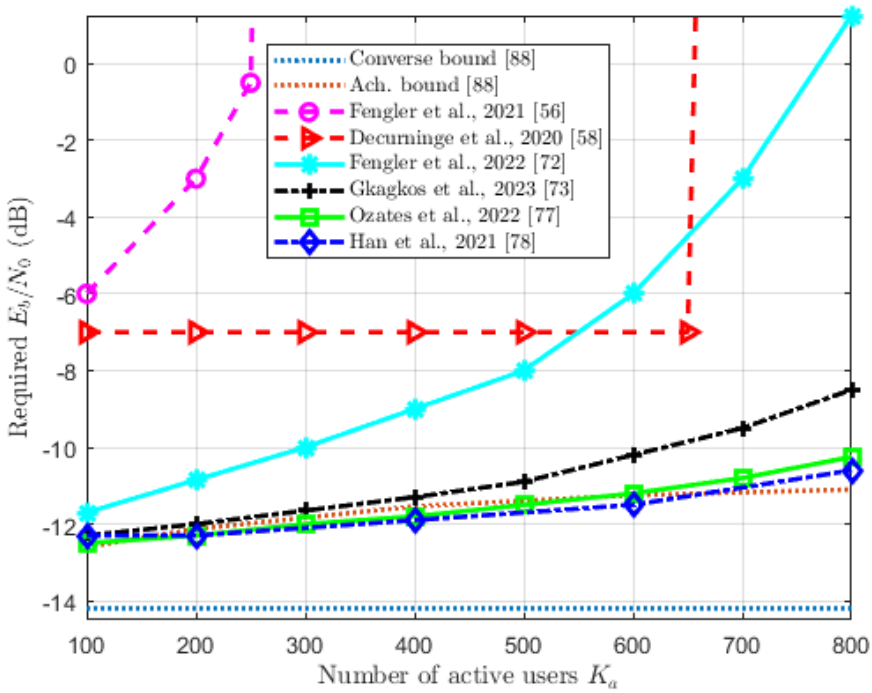


Figure 2.10: The required $E_b/N_0$ to achieve the target PUPE of 0.05 as a function of the number of active users for $n = 3200$, $B = 100$, and different coding schemes over MIMO Rayleigh fading channels.

In pilot-based algorithms, the channel estimation and data detection tasks are performed independently, hence the error from one block is propagated to the next one. To resolve this issue, there is also a new set of MIMO URA schemes which employ multi-layer iterative Bayesian decoder to jointly estimate each user's transmitted signal, and their corresponding channel coefficients [78, 79]. The superiority of Bayesian schemes over the pilot-based ones (see Figure

2.10) lies in their ability to carry out channel estimation and data detection in an iterative manner, allowing errors from one block to be resolved in a subsequent block. A summary of MIMO URA schemes is provided in Table 2.3.

| Paper | Contributions | Description |
|---|---|---|
| Gao et al., 2023 [88] | Bounds | Achievability and converse bounds. |
| Fengler et al., 2021 [56] | Coding scheme: CCS-based | The pilots at each slot are found using a newly proposed coordinate-wise activity detector. Then, the detected pilots of different slots are stitched together using parity bits. |
| Shyianov et al., 2021 [57] | Coding scheme: CCS-based | Instead of employing parity-check bits (as in [56]), the correlation between different slots is used for stitching pilots, hence, the energy efficiency is improved. |
| Decurninge et al., 2020 [58] | Coding scheme: tensor-based | Users' signal separation and detection is done by tensor decomposition. |
| Luan et al., 2022 [59] | Coding scheme: Tensor-based | A new tensor-based modulation scheme in two versions: one employing QPSK encoding and the other implementing Grassmann encoding. It shows improved energy efficiency over the tensor-based scheme in [58]. |
| Shao et al., 2022 [34] | Coding scheme: tensor-based | Users communicate with the BS through the RIS. For detecting the transmitted signals, a Bayesian approach is employed. |
| Fengler et al., 2022 [72] | Coding scheme: pilot-based | A pilot and a polar codeword are transmitted sequentially. For decoding, the active pilots and corresponding channel coefficients are detected using MMV-AMP and MMSE, respectively, and a polar decoder is employed for data detection. |
| Gkagkos et al., 2023 [73] | Coding scheme: pilot-based | A pilot and a random spread polar codeword are transmitted sequentially. An energy detector, an MMSE, and a polar list decoder are employed for pilot detection, channel estimation, and decoding, respectively. Finally, an SIC removes the contribution of successfully decoded signals. |

| | | |
|---|---|---|
| Ustinova et al., 2022 [74] | Coding scheme: pilot-based | OMP is employed for pilot detection and channel estimation, polar code is used for decoding, and SIC is used for reducing the interference. |
| Su et al., 2023 [75] | Coding scheme: pilot-based | A pilot is transmitted along with an LDPC coded and modulated codeword. At the receiver side, AMP is used for pilot detection and channel estimation, and an LDPC decoder is used for data detection. The selected pilot specifies the indexes over which every symbol of the LDPC codeword is transmitted. |
| Su et al., 2023 [76] | Coding scheme: pilot-based | A pilot and an LDPC codeword are spread and transmitted consecutively. At the receiver's end, an ML algorithm detects active pilots, an AMP technique estimates the corresponding channels, and an SIC block is applied following the LDPC decoder. |
| Ozates et al., 2023 [77] | Coding scheme: pilot-based | The channel length is slotted, and each user randomly selects a slot to transmit a pilot along with a polar codeword. For decoding, a generalized OMP and an MMSE estimator are used for pilot detection and channel estimation, the polar decoder is used for data detection, and SIC removes the contribution of the decoded messages. |
| Han et al., 2021 [78] | Coding scheme: Bayesian approach | Employs FEC code, and sparse spreading. |
| Jiang et al., 2023 [79] | Coding scheme: Bayesian approach | Combines sparse regression codes for encoding and a message-passing algorithm for decoding. |

Table 2.3: Summary of available coding schemes in URA with MIMO fading channels.

## 2.5 Chapter Summary

In this chapter, we examined unsourced random access as a highly effective scheme in massive machine-type communication. Initially, our focus was on investigating various multiple access techniques. Following that, we presented distinct advantages that URA exhibits in comparison to other multiple access methods, particularly in scenarios involving massive access. Subsequently, we delved into the challenges associated with URA and potential solutions. Lastly, we provided an overview of the literature on URA, placing emphasis on GMAC and fading channels. We reviewed research that introduced coding schemes and investigated performance bounds within the URA framework.

# Chapter 3

# Random Spreading With Power Diversity for URA over GMAC

In this chapter, we propose an extension of the random spreading idea in [64] by allowing for the use of different power levels among active users. In our approach, each user initially employs a polar code to encode their message, after which the encoded bits are spread using a randomly generated spreading sequence. This novel method involves categorizing active users into distinct groups and applying varying power levels to each group, ensuring compliance with the average power constraint. To achieve this, we introduce an optimization problem that determines the optimal number of groups, the expected number of users within each group, and the respective power levels. we will demonstrate the superior performance of our approach, particularly when dealing with a large number of active users.

The chapter is organized as follows. In Section 3.1, we describe the system model. In Section 3.2, the proposed random access scheme is described. In Section 3.3 an optimization problem for power allocation (PA) is formulated and solved. Simulation results are provided in Section 3.4, and our conclusions are given in Section 3.5.

## 3.1 System Model

We consider an unsourced random access model in which $K_a$ out of $K_T$ users are active at a given frame. The number of active users $K_a$ is assumed to be known at the receiver. Each active user transmits $B$ bits of information through $n$ channel uses. Assuming a GMAC, the received signal vector in the absence of synchronization errors and fading is written as

$$\mathbf{y} = \sum_{i=1}^{K_T} s_i \mathbf{x}_i(\mathbf{u}_i) + \mathbf{z}, \tag{3.1}$$

where $s_i$ is 1 for active users and 0 otherwise, $\mathbf{x}_i$ is the length $n$ spread polar-coded signal corresponding to the message bit sequence $\mathbf{u}_i \in \{0,1\}^B$ of user $i$, and $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ is the additive white Gaussian noise. Each user selects its message index uniformly from the set $\{i \in \mathbb{Z} : 1 \leq i \leq 2^B\}$. The average power of each user per channel use is set to $P$. Therefore, the energy-per-bit of the system can be written as

$$\frac{E_b}{N_0} = \frac{nP}{2B}, \tag{3.2}$$

and the PUPE of the system is defined as

$$P_e = \max_{\sum_{i=1}^{K_T} s_i = K_a} \frac{1}{K_a} \sum_{i=1}^{K_T} s_i \Pr(\mathbf{u}_i \notin \mathcal{L}(\mathbf{y})), \tag{3.3}$$

where $\mathcal{L}(\mathbf{y})$ is the list of decoded messages with size at most $K_a$. The primary purpose is to design encoding and decoding schemes to reach a PUPE less than the target block error probability $\epsilon$ with the lowest $E_b/N_0$.

## 3.2 Proposed Unsourced MAC Scheme

### 3.2.1 Encoder

The message selected by each user is divided into two parts with $B_s$ and $B_c = B - B_s$ bits, i.e., $\mathbf{u}_i = (\mathbf{u}_{is}, \mathbf{u}_{ic})$. The first part is used to map the preamble bits,

$\mathbf{u}_{is}$, to columns of a signature codebook $\mathbf{A} \in \mathbb{R}^{n_s, 2^{B_s}}$, where $n_s$ is the spreading sequence length. Since

$$P[\exists j \neq i : \mathbf{u}_{is} = \mathbf{u}_{js}] \leq \frac{K_a - 1}{2^{B_s}}, \tag{3.4}$$

$B_s$ is selected to make the right-hand side of (3.4) in such a way to satisfy the PUPE requirement. The elements of the matrix $\mathbf{A}$ are generated by first picking independent zero-mean Gaussian random variables with unit variance, forcing each column to have an average of zero by subtracting its mean from each element, and then scaling them to distinct power levels (the determination of the power levels will be discussed in the following section). Let us denote the column picked by the $i$th user by $\mathbf{a}_i$. The second part of the message, $\mathbf{u}_{ic}$, is encoded using a polar code. As in [64], this message is appended by $r$ CRC bits, and the result is passed to an $(n_c, B_c + r)$ polar encoder. The CRC bits are used to check the success of polar decoding. The polar codeword is modulated using binary shift keying (BPSK), resulting in $\mathbf{v}_i \in \{\pm 1\}^{n_c \times 1}$. The transmitted signal for user $i$ is then obtained as

$$\mathbf{x}_i = \mathbf{v}_i \otimes \mathbf{a}_i, \tag{3.5}$$

where $\otimes$ represents the Kronecker product.

### 3.2.2 Design of the Codebook

While the elements of the codebook $\mathbf{A}$ are selected in a similar fashion to [64], unlike the codebook in [64] where the empirical variance of all the columns of $\mathbf{A}$ are normalized, we divide the columns of $\mathbf{A}$ into $m$ groups, and assign different power levels to each. There are $l_k$ columns with power level $P_k$ in the $k$th group, with $k = 1, 2, \ldots, m$. Hence, with a uniform selection, the probability of choosing a column with power $P_k$ from the codebook is

$$\Pr(\|\mathbf{a}_i\|^2 = P_k) = \frac{l_k}{2^{B_s}}. \tag{3.6}$$

Since $K_a$ active users pick columns of $\mathbf{A}$ randomly, the number of users with power $P_k$ can be approximated by

$$K_k \approx \frac{l_k}{2^{B_s}} K_a, \tag{3.7}$$

for $K_a \gg 1$. Optimization of the codebook parameters $l_k$ and $P_k$ is described in the next section.

### 3.2.3 Decoder

The decoder is composed of three parts. A newly proposed covariance-based detector identifies the set of spreading sequences employed by the active users; a MMSE estimator is used to produce the soft estimates corresponding to the detected sequences; and finally, the estimated symbols are fed as input to the channel decoder, implemented as a polar list decoder. SIC is employed to remove the contribution of the successfully decoded messages from the received signal at each step, and the procedure is repeated until messages of all the active users are decoded or there are no successfully decoded users in an iteration.

#### 3.2.3.1 Covariance-based Detector

As an alternative to the energy detector in [64], which requires searching through possible codeword sequences, we propose an approach which generates an estimate of the covariance matrix of the received signal, and declares the signatures corresponding to the largest diagonal entries as the active ones. Namely, we write the remaining signal after last SIC step, $\mathbf{Y}' \in \mathbb{R}^{n_s \times n_c}$, as

$$\mathbf{Y}' = \mathbf{A}_a \mathbf{V} + \mathbf{Z}, \tag{3.8}$$

where $\mathbf{A}_a \in \mathbb{R}^{n_s \times K}$ and $\mathbf{V} \in \{\pm 1\}^{K \times n_c}$ are constructed by aggregating the signatures and codewords of the remaining $K$ active users; and, $\mathbf{Z} \in \mathbb{R}^{n_s \times n_c}$ is the noise matrix with independent and identically distributed zero-mean unit-variance Gaussian random variables as its elements. Note that since via a CRC check, we only remove the contribution from the correct decoded message at the SIC steps, (3.8) holds for all the iterations (with the remaining users' messages). We form

$$\mathbf{C} = \mathbf{A}_N^T \mathbf{Y}' \mathbf{Y}'^T \mathbf{A}_N, \tag{3.9}$$

where $\mathbf{A}_N$ is obtained by scaling the columns of the codebook to 1 (after removing signatures of correctly decoded users) to prevent higher detection probability for the sequences with greater power levels. The covariance-based detector outputs $K_\gamma = K_r + K_\delta$ signatures corresponding to largest diagonal elements of $\mathbf{C}$, where $K_r$ is the number of remaining users up to the current iteration and $K_\delta$ is a small integer.

Let us comment on the complexity of the proposed approach. The number of multiplications required to calculate the diagonal entries of matrix $\mathbf{C}$ in (3.9) is $N_c = n_s^2 n_c + 2^{B_s} n_s^2 + 2^{B_s} n_s$, while the number of multiplications required in the energy detector [64] is $N_e = n_s n_c 2^{(g+B_s)}$. Here $g$ denotes the length of each partition in the energy detector approach. Clearly, for moderate values of $g$, e.g., for $g = 8$, the complexity of the energy detector is significantly higher than that of the newly proposed one. For small $g$ values, however, the complexities of the two approaches are in the same order.

#### 3.2.3.2 Channel Decoder and SIC

Accumulating the signatures declared by the covariance-based detector in the matrix $\hat{\mathbf{A}}_{\mathcal{D}} \in \mathbb{R}^{n_s \times (K_r + K_\delta)}$, the MMSE estimate of $\mathbf{V}$ is obtained by [64]

$$\hat{\mathbf{V}} = \hat{\mathbf{A}}_{\mathcal{D}}^T \hat{\mathbf{C}}_y^{-1} \mathbf{Y}', \tag{3.10}$$

with $\hat{\mathbf{C}}_y = (\mathbf{I}_{n_s} + \hat{\mathbf{A}}_{\mathcal{D}} \hat{\mathbf{A}}_{\mathcal{D}}^T)$, where the $i$th row of $\hat{\mathbf{V}}$ (denoted by $\hat{\mathbf{v}}_i$) is the estimated codeword of user $i$. We treat $\hat{\mathbf{v}}_i$ as the output of an AWGN channel with noise variance $\hat{\sigma}_i^2$, where $\hat{\sigma}_i^2$ is obtained by picking the $i$th diagonal entry of the mean square error (MSE) matrix, MSE $= \mathbf{I} - \hat{\mathbf{A}}_{\mathcal{D}}^T \mathbf{C}_y^{-1} \hat{\mathbf{A}}_{\mathcal{D}}$. The $i$th user's message is then decoded by feeding $\mathbf{d}_i = 2\hat{\mathbf{v}}_i / \hat{\sigma}_i^2$ as the set of LLRs to the list decoder.

After polar decoding, the successfully decoded codewords that satisfy the CRC check are removed from the received signal. The residual received signal is then passed back to the covariance-based detector for the next iteration. This procedure is repeated until there are no successful decoding results or all the active users are decoded.

We finally note that there could be some false signatures declared by the covariance-based detector. In practice, during the decoding process, these falsely

detected signatures with higher power levels can adversely affect the decoding performance of the users with lower power levels. To alleviate this effect, we run the decoder $m+1$ times with indices $i = 0, 1, ..., m$. At the $i$th step, the signatures of the $i$ groups with the larger power levels are removed from the codebook when using the covariance-based detector. Therefore, the matrix $\hat{\mathbf{A}}_{\mathcal{D}}$ does not have the spreading sequences with high power levels in it. This helps the MMSE estimator in (3.10) to avoid possible negative effects of the falsely detected signatures with high power levels particularly when most of the users with these power levels are removed from the received signal by SIC. Note again that the falsely detected signatures do not negatively affect the SIC process since CRC check is used to verify the correctness of decoder outputs. The details of the decoding procedure are given as a pseudo-code in Algorithm 1.

**Algorithm 1:** Pseudo-code for the decoder steps.

```
Input: Y, A, and K_a.
flag = 1.
K_r = K_a.
Y' = Y.
while flag = 1 do  Decoding
    K_x = 0.
    for i = 0, 1, ..., m do
        i groups with higher powers are removed from A_N.
        S = ∅.
        Â_D is the output of covariance-based detector.
        Calculate V̂ = Â_D^T Ĉ_y^{-1} Y'.
        Feed d_i = 2v̂_i/σ̂_i^2 to the list decoder.
        update S as the set of decoded codewords.
        K_r = K_r − |S| and K_x = K_x + |S|.
        Y' = Y' − A_S V_S.
        if K_x ≥ 1 then
            break.
        end
    end
    if K_x = 0 or K_r = 0 then
        flag = 0.
    end
end
```

## 3.3 Design of Spreading Sequences: Optimal Power Allocation

### 3.3.1 Selection of Optimal Parameters

Suppose that active users are divided into $m$ groups consisting of $K_1, K_2, ..., K_m$ users with power levels $P_1 \leq P_2 \leq ... \leq P_m$, where $K_1 + K_2 + ... + K_m = K_a$. Note that during the interference cancellation procedure, the users wih the highest power levels are expected to be decoded first. Focusing on (3.5), we notice that $\mathbf{x}_i$ is obtained by normalizing a zero-mean Gaussian vector, hence we model the aggregate interference as a Gaussian random vector added to each user's signal. For instance, for the group with the highest power level, $K_m$ users with power $P_m$ experience interference with variance $\sigma_m^2 = 1 + K_1P_1 + K_2P_2 + ... + K_{m-1}P_{m-1}$. After successful decoding and interference cancellation, we decode the $K_{m-1}$ users with power level $P_{m-1}$, which experience interference with variance $\sigma_{m-1}^2 = 1 + K_1P_1 + K_2P_2 + ... + K_{m-2}P_{m-2}$. Similarly, for the $j$th group, the $K_j$ users with power $P_j$ experience interference with variance $\sigma_j^2 = 1 + \sum_{i=1}^{j-1} K_iP_i$.

**Corollary 1.** *The minimum required power for the $j$th group becomes*

$$P_j = \frac{\alpha_{\min}(K_j)}{1 - (K_j - 1)\alpha_{\min}(K_j)} \left(1 + \sum_{i=1}^{j-1} K_iP_i\right), \tag{3.11}$$

*where $\alpha_{\min}(K_j)$ is the minimum required SINR for achieving a target PUPE in a group with $K_j$ users.*

*Proof.* Since we are treating interference as noise, from the perspective of a user, the SINR becomes the important performance metric. In a group with $K_j$ users with power level $P_j$, and noise variance $\sigma_j^2$, we need

$$\alpha_{\min}(K_j) \leq \frac{P_j}{\sigma_j^2 + (K_j - 1)P_j}. \tag{3.12}$$

Since our objective is to minimize the total power, we can select the power level for each user as

$$P_j = \frac{\alpha_{\min}(K_j)}{1-(K_j-1)\alpha_{\min}(K_j)}\sigma_j^2. \tag{3.13}$$

□

We can rewrite (3.11) as

$$P_j = \gamma_j \prod_{i=1}^{j-1}(1+K_i\gamma_i)\ \forall j = 1,2,\ldots,m, \tag{3.14}$$

where $\gamma_j = \dfrac{\alpha_{\min}(K_j)}{1-(K_j-1)\alpha_{\min}(K_j)}$. Note that $\alpha_{\min}(K_j)$ depends on the particular coding and transmission scheme employed, and can be determined via simulations. Therefore, the total power can be written as a function of $K_1, K_2, ..., K_m$ as

$$\begin{aligned} P_T &= K_1P_1 + K_2P_2 + ... + K_mP_m \\ &= \prod_{i=1}^{m}(1+K_i\gamma_i) - 1. \end{aligned} \tag{3.15}$$

We can find values of $K_1, K_2, ..., K_m$ which minimize the total power by solving the following optimization problem:

$$(\hat{K}_1, ..., \hat{K}_m) = \underset{K_1,...,K_m}{\arg\min}\, P_T, \quad \text{s.t.} \sum_{i=1}^{m} K_i = K_a. \tag{3.16}$$

To proceed further, we relax the integer constraint on $K_j$'s, and use the method of Lagrange multipliers. The cost function to be minimized is

$$\mathbf{J}(K_1,\ldots,K_m) = \prod_{i=1}^{m}(1+K_i\gamma_i) + \lambda\left(\sum_{i=1}^{m} K_i - K_a\right) \tag{3.17}$$

where $\lambda$ is the Lagrange multiplier. Setting the derivative of the Lagrangian function to zero gives

$$\frac{K_1\gamma_1' + \gamma_1}{1+K_1\gamma_1} = \frac{K_2\gamma_2' + \gamma_2}{1+K_2\gamma_2} = ... = \frac{K_m\gamma_m' + \gamma_m}{1+K_m\gamma_m} = -\frac{\lambda}{P_T+1}. \tag{3.18}$$

Strictly speaking, while $\alpha_{\min}(K_j)$ is a function of $K_j$, this value is almost a constant (for a given target PUPE level). Dropping this dependence, after some calculations, we obtain

$$\frac{K_i\gamma_i' + \gamma_i}{1+K_i\gamma_i} = \frac{\alpha_{\min}}{1-(K_i-1)\alpha_{\min}}. \tag{3.19}$$

Since $\dfrac{\alpha_{\min}}{1-(K_i-1)\alpha_{\min}}$ is a one-to-one function of $K_i$, the optimal values of $K_i$ satisfy $K_1 = K_2 = ... = K_m = K_0$, with $K_0 = K_a/m$. We can also argue that (with a constant $\alpha_{\min}$), the Hessian for $\mathbf{J}(K_1, \ldots, K_m)$ is given as

$$\nabla^2(\mathbf{J}) = (P_T+1)\text{diag}\left(\frac{K_1\gamma_1'' + 2\gamma_1'}{1+K_1\gamma_1}, \ldots, \frac{K_m\gamma_m'' + 2\gamma_m'}{1+K_m\gamma_m}\right) \tag{3.20}$$

with

$$\gamma_i' = \frac{\partial \gamma_i}{\partial K_i} = \left(\frac{\alpha_{\min}}{1-(K_i-1)\alpha_{\min}}\right)^2, \tag{3.21}$$

$$\gamma_i'' = \frac{\partial^2 \gamma_i}{\partial K_i^2} = 2\left(\frac{\alpha_{\min}}{1-(K_i-1)\alpha_{\min}}\right)^3, \tag{3.22}$$

where $\text{diag}(\zeta_1, ..., \zeta_l)$ is an $l \times l$ diagonal matrix with the $i$th diagonal element $\zeta_i$. From (3.13), we observe that $\dfrac{\alpha_{\min}}{1-(K_i-1)\alpha_{\min}}$ is proportional to the minimum required power, which is always positive. Thus, the Hessian in (3.20) is a positive definite matrix, and (3.16) is verified to be a convex optimization problem.

Determining that the number of users in all the groups should be the same, we optimize the number of groups via

$$m = \min_{\hat{m}} \left(1 + \frac{K_a\gamma_0}{\hat{m}}\right)^{\hat{m}}. \tag{3.23}$$

### 3.3.2 General case

We highlight that, even though the optimal PA approach is applied to the random spreading scheme, the basic ideas can be extended to other unsourced MAC scenarios (for which the interference is treated as noise) as well by solving (3.23) for the specific $\alpha_{\min}(K)$ values.

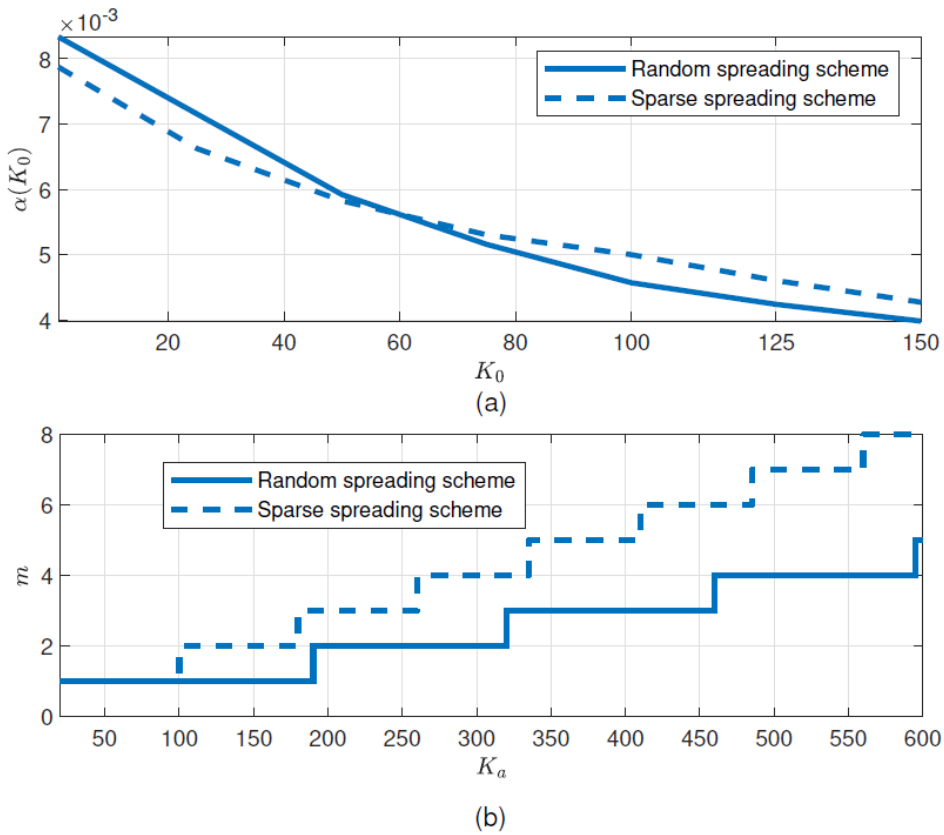


Figure 3.1: (a) The required SINR and (b) optimal number of groups for random spreading scheme with $n_s = 117$ and $n_c = 256$ and sparse spreading scheme in [89].

## 3.4 Numerical Results

In this section, we study the performance of the proposed approach and compare it with other existing solutions for unsourced MAC. The total number of channel uses is selected as $n \cong 30000$ ($n_s = 117$ and $n_c = 256$), the number of active users is $150 \leq K_a \leq 600$, the list size for the polar decoder is $L = 512$, and the target PUPE is set to $P_e = 0.05$. The number of information bits per message is chosen as $B = 100$, where $B_s$ is selected as $B_s = 14$ for $150 \leq K_a < 200$, $B_s = 15$ for $200 \leq K_a \leq 250$, $B_s = 16$ for $250 < K_a \leq 350$, $B_s = 17$ for $350 < K_a \leq 500$, and $B_s = 18$ for $500 < K_a \leq 600$. Note that, the probability that three or more

users are in collision is very small for the selected values of $B_s$, hence their effects can be neglected. Also, as discussed in [64], the random spreading scheme with channel coding rate lower than 1/2 is able to resolve the possible collision of two users, providing a good performance.

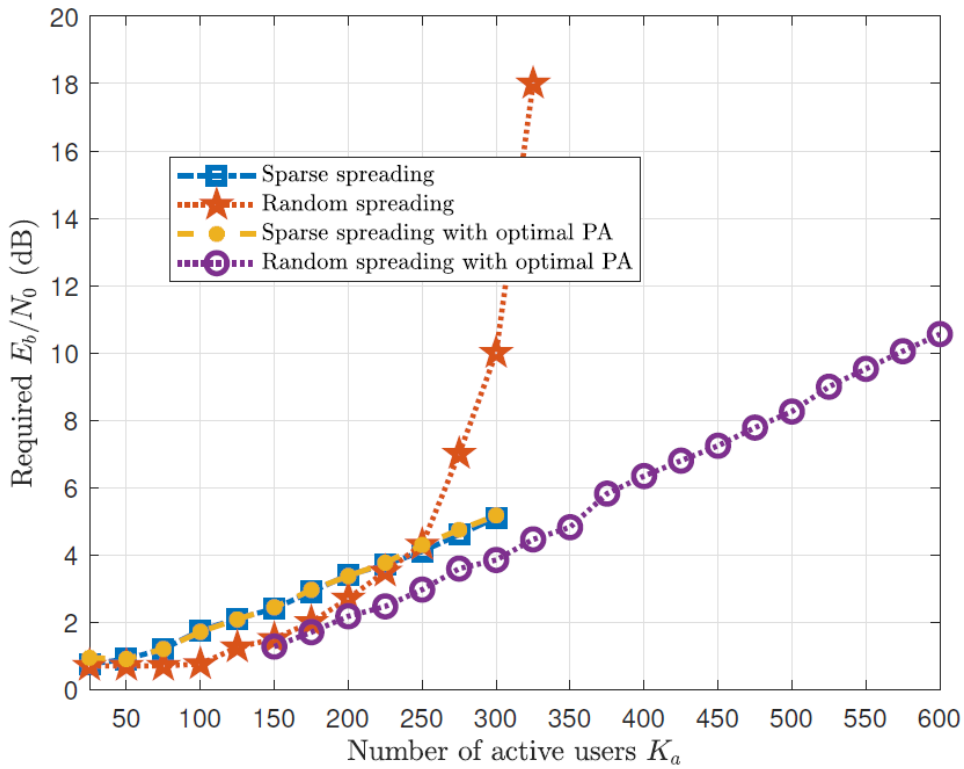


Figure 3.2: The required $E_b/N_0$ as a function of the number of active users in the sparse spreading scheme [89], the random spreading scheme [64], the sparse spreading with optimal PA, and the random spreading with optimal PA.

To obtain $\alpha_{\min}(K_0)$ for the random spreading scheme, the method in [64] is run for different values of $K_0$, and the result is shown in Figure 3.1(a) for $B_s = 14$ and 18. One should keep in mind that the required SINR must be computed for different values of $B_s$ separately, because in the random spreading scheme, it is sensitive to $B_s$. Plugging $\alpha_{\min}(K_0)$ into (3.23), the optimal number of groups, $m$, is obtained (as depicted in Figure 3.1(b)). Clearly, for $K_a \leq 125$, only one group should be employed, however, beyond that number of users, employing more than one group with different power levels becomes optimal.

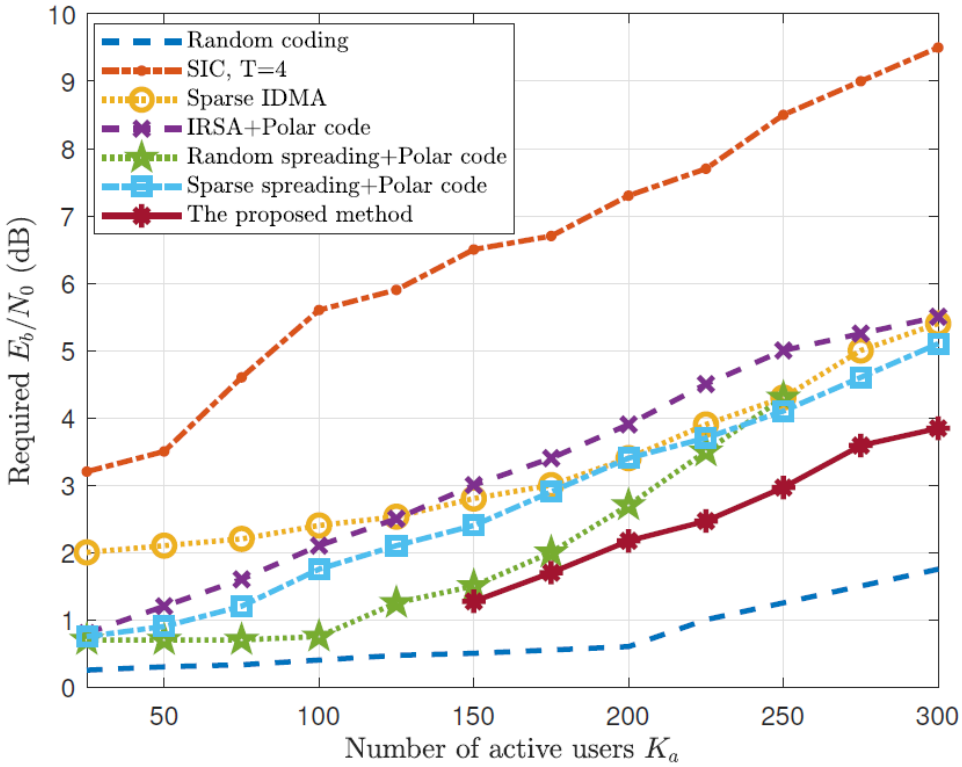


Figure 3.3: The required $E_b/N_0$ as a function of the number of active users for the proposed scheme, the random coding bound in [64], SIC-based scheme with $T = 4$ [60], sparse IDMA [62], IRSA scheme [67], random spreading scheme [64], and sparse spreading scheme [89].

We also apply our approach to the sparse spreading idea of [89] to determine the optimal number of groups and the corresponding power levels without any need for extensive simulations. The results are depicted in Figure 3.1. We compare the required $E_b/N_0$ of the scheme in [89], random spreading scheme in [64], sparse spreading with optimal PA, and random spreading with optimal PA in Figure 3.2. It can be inferred from this result that applying optimal PA on the random spreading scheme decreases the required $E_b/N_0$ of the system remarkably, especially, for the larger number of active users, i.e., for $K_a > 225$. Moreover, it is shown that the required $E_b/N_0$ is the same for the sparse spreading scheme with optimal PA and sparse spreading scheme in [89] where the parameters are empirically chosen.

In Figure 3.3, the performance of the proposed random spreading solution with optimal PA is compared with the other existing unsourced MAC schemes. The result clearly indicates that the proposed method offering superior performance, particularly, when the number of active users is large.

We further note our observation that the detection probability of the covariance-based detector is slightly higher than that of the energy detector for manageable values of $g$, and for the specific parameters of $n_s$, $n_c$, and $B_s$ in this section. Regarding its complexity, for $K_a = 500$ and $B_s = 17$, the total number of multiplications required by the proposed covariance-based detector and the energy detector with $g = 1$ are $N_c = 1.81 \times 10^9$ and $N_e = 7.85 \times 10^9$, respectively. That is, the computational complexity of the covariance-based detector and energy detector with $g = 1$ are comparable.

## 3.5 Chapter Summary

In this chapter, we have studied unsourced MAC with polar codes and random spreading, and developed an approach which divides the active users into different groups with varying transmit power levels. The optimal number of groups and the power levels are selected through a suitably formulated optimization problem, and it is shown by numerical evaluations that the idea of using different power levels provides a significant reduction in the required $E_b/N_0$, particularly, for systems with a large number of active users.

# Chapter 4

# MS-MRA: Multi-stage URA Set-up over MIMO Fading Channels

In this chapter, we study the problem of URA over Rayleigh block-fading channels with a receiver equipped with multiple antennas. We propose a slotted structure with multiple stages of orthogonal pilots, each of which is randomly picked from a codebook. In the proposed signaling structure, each user encodes its message using a polar code and appends it to the selected pilot sequences to construct its transmitted signal. Accordingly, the transmitted signal is composed of multiple orthogonal pilot parts and a polar-coded part, which is sent through a randomly selected slot. The performance of the proposed scheme is further improved by randomly dividing users into different groups each having a unique interleaver-power pair. We also apply the idea of multiple stages of orthogonal pilots to the case of a single receive antenna. In all the set-ups, we use an iterative approach for decoding the transmitted messages along with a suitable successive interference cancellation technique. The use of orthogonal pilots and the slotted structure lead to improved accuracy and reduced computational complexity in the proposed set-ups, and make the implementation with short blocklengths more viable. Performance of the proposed set-ups is illustrated via extensive simulation results which show that the proposed set-ups with multiple antennas perform better than the existing MIMO URA solutions for both short and large

blocklengths, and that the proposed single-antenna set-ups are superior to the existing single-antenna URA schemes.

The chapter is organized as follows. In Section 4.1, an introductory overview for the chapter is presented. Section 4.2 presents the system model for the proposed framework. The encoding and decoding procedures of the proposed schemes are introduced in Section 4.3. In Section 4.4, extensive numerical results and examples are provided. Finally, Section 4.5 provides our conclusions.

## 4.1 Introduction

Most coding schemes in URA employ non-orthogonal pilots/sequences for identification and estimation purposes [36, 56, 64, 65, 72, 73]. Performance of detectors and channel estimators may be improved in terms of accuracy and computational complexity by employing a codebook of orthogonal pilots; however, this significantly increases the amount of collisions due to the limited number of available orthogonal pilot sequences. To address this problem, the proposed schemes in this chapter employ multiple stages of orthogonal pilots combined with an iterative detector.

In the proposed scheme, the transmitted signal of each user is composed of $J+1$ stages: a polar codeword appended to $J$ independently generated orthogonal pilots. Thus, the scheme is called multi-stage set-up with multiple receive antennas (MS-MRA). At each iteration of MS-MRA at the receiver side, only one of the pilot parts is employed for pilot detection and channel estimation, and the polar codeword is decoded using a polar list decoder. Therefore, the transmitted pilots in the remaining $J-1$ pilot parts are still unknown. To determine the active pilots in these, we adopt two approaches. In the first one, all the pilot bits are coded jointly with the data bits and CRC bits (therefore, the transmitted bits of all the pilot parts are detected after successful polar decoding). As a second approach, to avoid waste of resources, we propose an enhanced version of the MS-MRA, where only data and CRC bits are fed to the polar encoder. At the receiver

side, the decoder iteratively moves through different $J+1$ parts of the signal to detect all the parts of an active user's message. Since it does not encode the pilot bits, this is called MS-MRA without pilot bits encoding (MS-MRA-WOPBE). We further improve the performance of the MS-MRA by randomly dividing users into different groups. In this scheme (called multi-stage set-up with user grouping for multiple receive antennas (MSUG-MRA)), each group is assigned a unique interleaver-power pair. Transmission with different power levels increases the decoding probability of the users with the highest power (because they are perturbed by interfering users with low power levels). Since successfully decoded signals are removed using SIC, users with lower power levels have increased chance of being decoded in the subsequent steps. By repeating each user's signal multiple times, we further extend the idea in MS-MRA and MSUG-MRA to the case of a single receive antenna. These extensions are called multi-stage set-up with a single receive antenna (MS-SRA) and multi-stage set-up with user grouping for a single receive antenna (MSUG-SRA).

Several studies have investigated Rayleigh block-fading channels in a massive MIMO setting [37, 56, 72, 73]. In [56], a covariance-based activity detection (AD) algorithm is used to detect the active messages. A pilot-based scheme is introduced in [72] where non-orthogonal pilots are employed for detection and channel estimation, and a polar list decoder is used for decoding messages. Furthermore, in a scheme called FASURA [73], each user transmits a signal containing a non-orthogonal pilot and a randomly spread polar code. The AD algorithm in [56] performs well in the fast fading scenario (e.g., when $L_c \leq 320$); however, it is not implementable with larger blocklengths due to run-time complexity scaling with $L_c^2$. In contrast, the schemes in [72, 73] work well in the large-blocklength regimes (e.g., for $L_c = 3200$); that is, in a slow fading environment where large blocklengths can be employed, their decoding performance is better than that of [56]. We demonstrate that, while the covariance-based AD algorithm in [56] suffers from performance degradation with large blocklengths, and the algorithms in [72, 73] do not work well in the short blocklength regime (hence not suitable for fast fading scenarios), the proposed schemes in this chapter have a superior performance in both regimes.

Our contributions are summarized as follows:

- We propose a URA set-up with multiple receive antennas, namely MS-MRA. The proposed set-up offers comparable performance with the existing schemes with large blocklengths, while having lower computational complexity. Moreover, for the short-blocklength scenario, it significantly improves the state-of-the-art.
- We provide a theoretical analysis to predict the error probability of the MS-MRA, taking into account all the sources of error, namely, errors resulting from pilot detection, channel estimation, channel decoding, SIC, and collisions.
- We extend the MS-MRA set-up by randomly dividing the users into groups, i.e., MSUG-MRA, which is more energy-efficient than MS-MRA and other MIMO URA schemes.
- Two URA set-ups with a single receive antenna, called MS-SRA and MSUG-SRA, are provided by adopting the ideas of the MS-MRA and MSUG-MRA to the case of a single receive antenna. They perform better than the alternative solutions over fading channels.

## 4.2 System Model

We consider an unsourced random access model over a block-fading wireless channel. The BS is equipped with $M$ receive antennas connected to $K_T$ potential users, for which $K_a$ of them are active in a given frame. Assuming that the channel coherence time is larger than $L$, we divide the length-$n$ time-frame into $S$ slots of length $L$ each ($n = SL$). Each active user randomly selects a single slot to transmit $B$ bits of information. In the absence of synchronization errors, the received signal vector corresponding to the $s$th slot at the $m$th antenna is written

as

$$\mathbf{y}_{m,s} = \sum_{i \in \mathcal{K}_s} h_{m,i} \mathbf{x}\left(\mathbf{w}(i)\right) + \mathbf{z}_{m,s}, \tag{4.1}$$

where $\mathbf{y}_{m,s} \in \mathbb{C}^{1\times L}$, $\mathcal{K}_s$ denotes the set of active user indices available in the $s$th slot, $K_s = |\mathcal{K}_s|$, $\mathbf{x}\left(\mathbf{w}(i)\right) \in \mathbb{C}^{1\times L}$ is the encoded and modulated signal corresponding to the message bit sequence $\mathbf{w}(i) \in \{0,1\}^B$ of the $i$th user, $h_{m,i} \sim \mathcal{CN}(0,1)$ is the channel coefficient between the $i$th user and the $m$th receive antenna, and $\mathbf{z}_{m,s} \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2 \mathbf{I}_L)$ is the circularly symmetric complex white Gaussian noise vector. Letting $\mathcal{K}_a$ and $\mathcal{L}_d$ be the set of active user indices and the list of decoded messages, respectively, the PUPE of the system is defined in terms of the probability of false-alarm, $p_{fa}$, and the probability of missed-detection, $p_{md}$, as

$$P_e = p_{fa} + p_{md}, \tag{4.2}$$

where $p_{md} = \dfrac{1}{K_a} \sum_{i \in \mathcal{K}_a} \Pr(\mathbf{w}(i) \notin \mathcal{L}_d)$ and $p_{fa} = \mathbb{E}\left\{\dfrac{n_{fa}}{|\mathcal{L}_d|}\right\}$, with $n_{fa}$ being the number of decoded messages that were indeed not sent. The energy-per-bit of the set-up can be written as $\dfrac{E_b}{N_0} = \dfrac{LP}{\sigma_z^2 B}$, where $P$ denotes the average power of each user per channel use. The objective is to minimize the required energy-per-bit for a target PUPE.

## 4.3 URA with Multiple Stages of Orthogonal Pilots

### 4.3.1 MS-MRA Encoder

In this part, we introduce a multi-stage signal structure which is used in both of the proposed URA set-ups. As shown in Figure 4.1, we divide the message of the $i$th user into $J+1$ parts (one coded part and $J$ pilot parts) denoted by $\mathbf{w}_c(i)$ and $\mathbf{w}_{p_j}(i), j = 1, 2, ..., J$ with lengths $B_c$ and $B_p$, respectively, where $B_c + JB_p = B$. The $i$th user obtains its $j$th pilot sequence, $\mathbf{b}_{ji}$, with length $n_p = 2^{B_p}$ by mapping

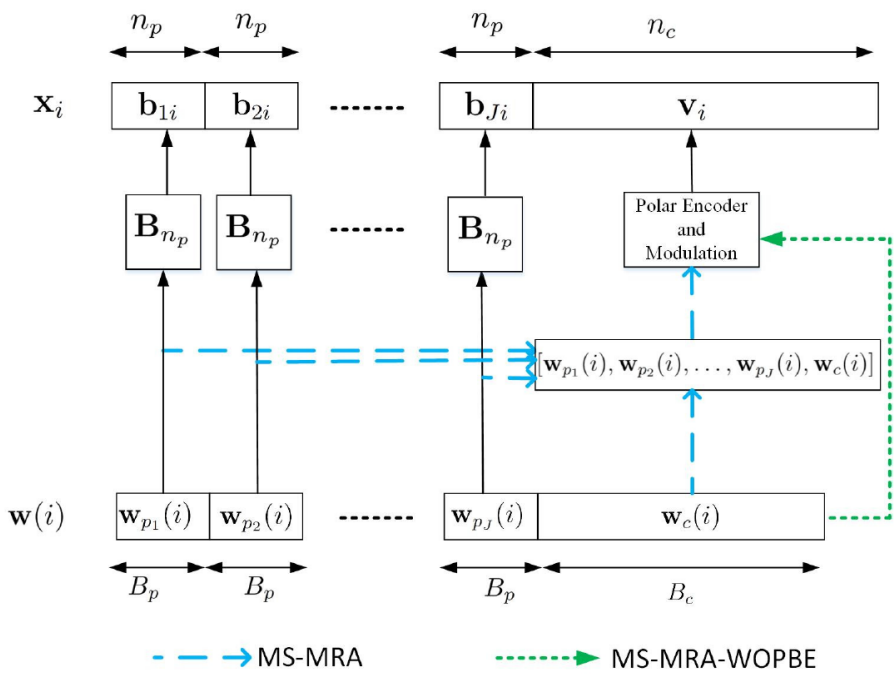


Figure 4.1: Illustration of the encoding process in the proposed MS-MRA schemes.

$\mathbf{w}_{p_j}(i)$ to the orthogonal rows of an $n_p \times n_p$ Hadamard matrix $\mathbf{B}_{n_p}$, which is generated as

$$\mathbf{B}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad \mathbf{B}_{2^i} = \mathbf{B}_2 \otimes \mathbf{B}_{2^{i-1}} \quad \forall \quad i = 2, 3, \ldots,$$

where $\otimes$ represents the Kronecker product. Since the number of possible pilots in the orthogonal Hadamard codebook is limited, it is likely that the users will be in collision in certain pilot segments, that is, they share the same pilots with the other users. However, the parameters are chosen such that two different users are in a complete collision in all the pilot parts with a very low probability. To construct the coded sequence of the $i$th user, we accumulate all the message bits in a row vector as

$$\mathbf{w}(i) = [\mathbf{w}_{p_1}(i), \mathbf{w}_{p_2}(i), \ldots, \mathbf{w}_{p_J}(i), \mathbf{w}_c(i)], \tag{4.3}$$

and pass it to an $(2n_c,\ B + r)$ polar code, where $r$ is the number of CRC bits. Note that contrary to the existing schemes in URA, we feed not only data bits but

the pilot bit sequences to the encoder. Hence, in the case of successful decoding, all the pilot sequences for the user can be retrieved. The polar codeword is then modulated using quadrature phase shift keying (QPSK), resulting in $\mathbf{v}_i \in \{\sqrt{P_c/2}(\pm 1 \pm j)\}^{1\times n_c}$, where $P_c$ is the average power of the polar coded part, and Gray mapping is used. The overall transmitted signal for the $i$th user consists of $J$ pilot parts and one coded part, i.e.,

$$\mathbf{x}_i = \left[\sqrt{P_p}\mathbf{b}_{1i}, \sqrt{P_p}\mathbf{b}_{2i}, \ldots, \sqrt{P_p}\mathbf{b}_{Ji}, \mathbf{v}_i\right] \in \mathbb{C}^{1\times L}, \tag{4.4}$$

where $L = n_c + Jn_p$ and $P_p$ denotes the average power of the pilot sequence. Accordingly, the received signal in a slot is composed of $J+1$ parts, for which, at each iteration, the decoding is done by employing one of the $J$ pilot parts (sequentially) and the coded part of the received signal. Generally, only the non-colliding users can be decoded. Some non-colliding users in the current pilot stage may experience collisions in the other pilot parts. Therefore, by successfully decoding and removing them using SIC, the collision density is reduced, and with further decoding iterations, the effects of such collisions are ameliorated.

### 4.3.2 MS-MRA Decoder

We now introduce the decoding steps of MS-MRA where the transmitted signal in (4.4) is received by $M$ antennas through a fading channel. The $j$th pilot part and the polar coded part of the received signal in the $s$th slot of the MS-MRA can be modeled using (4.1) as

$$\mathbf{Y}_{p_j} = \sqrt{P_p}\mathbf{H}\mathbf{B}_j + \mathbf{Z}_{p_j} \in \mathbb{C}^{M\times n_p},\ j = 1, 2, \ldots, J, \tag{4.5}$$

$$\mathbf{Y}_c = \mathbf{H}\mathbf{V} + \mathbf{Z}_c \in \mathbb{C}^{M\times n_c}, \tag{4.6}$$

where $\mathbf{H} \in \mathbb{C}^{M\times K_s}$ is the channel coefficient matrix with $h_{m,i}$ in its $m$th row and $i$th column, $\mathbf{Z}_{p_j}$ and $\mathbf{Z}_c$ consist of independent and identically distributed (i.i.d.) noise samples drawn from $\mathcal{CN}(0, \sigma_z^2)$ (i.e., a circularly symmetric complex Gaussian distribution), and $\mathbf{b}_{ji}$ and $\mathbf{v}_i$ determine the rows of $\mathbf{B}_j \in \{\pm 1\}^{K_s\times n_p}$ and $\mathbf{V} \in \{\sqrt{P_c/2}(\pm 1 \pm j)\}^{K_s\times n_c}$, respectively, with $i \in \mathcal{K}_s$. Note that we have removed the slot indices from the above matrices to simplify the notation.

The decoding process is comprised of five different steps that work in tandem. A pilot detector based on a Neyman-Pearson (NP) test identifies the active pilots in the current pilot part; channel coefficients corresponding to the detected pilots are estimated using a channel estimator; maximum-ratio combining (MRC) is used to produce a soft estimate of the modulated signal; after demodulation, the signal is passed to a polar list decoder; and, the successfully decoded codewords are added to the list of successfully decoded signals before being subtracted from the received signal via SIC. The process is repeated until there are no successfully decoded users in $J$ consecutive SIC iterations. In the following, $\mathbf{Y}'_{p_j}$ and $\mathbf{Y}'_c$ denote the received signals in (4.5) and (4.6) after removing the list of messages successfully decoded in the current slot up to the current iteration.

#### 4.3.2.1 Pilot Detection Based on NP Hypothesis Testing

At the $j$th pilot part, we can write the following binary hypothesis testing problem:

$$\begin{aligned} \mathbf{u}_{ji}|\mathcal{H}_0 &\sim \mathcal{CN}\left(\mathbf{0}, \sigma_z^2 \mathbf{I}_M\right), \\ \mathbf{u}_{ji}|\mathcal{H}_1 &\sim \mathcal{CN}\left(\mathbf{0}, \sigma_1^2 \mathbf{I}_M\right), \end{aligned} \tag{4.7}$$

where $\sigma_1 = \sqrt{\sigma_z^2 + m_{ij} n_p P_p}$, $\mathbf{u}_{ji} = \mathbf{Y}'_{p_j} \bar{\mathbf{b}}_i^H / \sqrt{n_p}$, with $\bar{\mathbf{b}}_i = \left[\mathbf{B}_{n_p}\right]_{(i,:)}$, $\mathcal{H}_1$ and $\mathcal{H}_0$ are alternative and null hypotheses that show the existence and absence of the pilot $\bar{\mathbf{b}}_i$ at the $j$th pilot part, respectively, and $m_{ij}$ is the number of users that pick the pilot $\bar{\mathbf{b}}_i$ as their $j$th pilots.

**Lemma 1.** *Let $\hat{\mathcal{D}}_j$ be the estimate of the set of active rows of $\mathbf{B}_{n_p}$ in the $j$th pilot part. Using a $\gamma-$level Neyman-Pearson hypothesis testing (where $\gamma$ is the bound on the false-alarm probability), $\hat{\mathcal{D}}_j$ can be obtained as*

$$\hat{\mathcal{D}}_j = \left\{ l : \mathbf{u}_{jl}^H \mathbf{u}_{jl} \geq \tau_0' \right\}, \tag{4.8}$$

*where $\tau_0' = 0.5\sigma_z^2 \Gamma_{2M}^{-1}(1-\gamma)$, $\Gamma_k(.)$ denotes the cumulative distribution function of the chi-squared distribution with $k$ degrees of freedom $\chi_k^2$, and $\Gamma_k^{-1}(.)$ is its inverse.*

*Proof.* The likelihood ratio for (4.7) is given by $L(\mathbf{u}_{ji}) = \dfrac{\mathbb{P}\left(\mathbf{u}_{ji}|\mathcal{H}_1\right)}{\mathbb{P}\left(\mathbf{u}_{ji}|\mathcal{H}_0\right)} = \dfrac{1}{\sigma_1^M} e^{\mathbf{u}_{ji}^H \mathbf{u}_{ji}/\sigma_0^2}$, where $\sigma_0^2 = \sigma_z^2 \left(\dfrac{\sigma_z^2 + m_{ij} n_p P_p}{m_{ij} n_p P_p}\right)$. Thus, the Neyman-Pearson test for detection of $\bar{\mathbf{b}}_i$ is obtained by

$$\begin{aligned}\delta_{NP}(\mathbf{u}_{ji}) &= \begin{cases} 1 & L(\mathbf{u}_{ji}) \geq \tau_0 \\ 0 & L(\mathbf{u}_{ji}) < \tau_0 \end{cases} \\ &= \begin{cases} 1 & \mathbf{u}_{ji}^H \mathbf{u}_{ji} \geq \tau_0' \\ 0 & \mathbf{u}_{ji}^H \mathbf{u}_{ji} < \tau_0' \end{cases}, \end{aligned} \tag{4.9}$$

where $\tau_0' = \sigma_0^2 \ln(\tau_0 \sigma_1^M)$. The false-alarm probability of the above decision rule is calculated as

$$\begin{aligned} P_F(\delta_{NP}) &= \mathbb{P}\left(\mathbf{u}_{ji}^H \mathbf{u}_{ji} \geq \tau_0' | \mathcal{H}_0\right) \\ &\overset{(a)}{=} 1 - \Gamma_{2M}\left(\frac{2}{\sigma_z^2}\tau_0'\right), \end{aligned} \tag{4.10}$$

where $(a)$ follows from the fact that $\dfrac{2}{\sigma_z^2}\mathbf{u}_{ji}^H \mathbf{u}_{ji}|\mathcal{H}_0 \sim \chi_{2M}^2$, with $\chi_k^2$ denoting the chi-squared distribution with $k$ degrees of freedom. To find the threshold for a $\gamma-$level Neyman-Pearson test, the probability of the false-alarm in (4.10) must satisfy $P_F(\delta_{NP}) \leq \gamma$. Therefore, the threshold in (4.9) is obtained as $\tau_0' = \dfrac{\sigma_z^2}{2}\Gamma_{2M}^{-1}(1-\gamma)$.

□

The detection probability of a non-colliding user ($m_{ij} = 1$) is then obtained as

$$\begin{aligned} P_D(\delta_{NP}) =& \mathbb{P}\left(\mathbf{u}_{ji}^H \mathbf{u}_{ji} \geq \tau_0' | \mathcal{H}_1\right) \\ \overset{(a)}{=}& 1 - \Gamma_{2M}\left(\frac{2\tau_0'}{\sigma_z^2 + n_p P_p}\right) \\ =& 1 - \Gamma_{2M}\left(\frac{\sigma_z^2 \Gamma_{2M}^{-1}(1-\gamma)}{\sigma_z^2 + n_p P_p}\right), \end{aligned} \tag{4.11}$$

where in (a), we use the fact that $\dfrac{2}{\sigma_1^2}\mathbf{u}_{ji}^H \mathbf{u}_{ji}|\mathcal{H}_1 \sim \chi_{2M}^2$. Note that a higher probability of detection is obtained in the general case of $m_{ij} > 1$. It is clear that the probability of detection is increased by increasing the parameters $\gamma$, $n_p$, $P_p$, and $M$.

#### 4.3.2.2 Channel Estimation

Let $\mathbf{B}_{\hat{\mathcal{D}}_j} \in \{\pm 1\}^{|\hat{\mathcal{D}}_j| \times n_p}$ be a sub-matrix of $\mathbf{B}_{n_p}$ consisting of the detected pilots in (4.8), and suppose that $\tilde{\mathbf{b}}_{jk} = \left[\mathbf{B}_{\hat{\mathcal{D}}_j}\right]_{(k,:)}$ is the corresponding pilot of the $i$th user. Since the rows of the codebook are orthogonal to each other, the channel coefficient vector of the $i$th user can be estimated as

$$\hat{\mathbf{h}}_i = \frac{1}{n_p\sqrt{P_p}} \mathbf{Y'}_{p_j} \tilde{\mathbf{b}}_{jk}^T. \tag{4.12}$$

If the $i$th user is in a collision ( $m_{ij} > 1$), (4.12) gives an unreliable estimate of the channel coefficient vector. However, this is not important since a CRC check is employed after decoding and such errors do not propagate.

#### 4.3.2.3 MRC, Demodulation, and Channel Decoding

Let $\mathbf{h}_i$ be the channel coefficient vector of the $i$th user, where $i \in \tilde{\mathcal{S}}_s$ with $\tilde{\mathcal{S}}_s$ denoting the set of remaining users in the $s$th slot. Using $\hat{\mathbf{h}}_i$ in (4.12), the modulated signal of the $i$th user can be estimated employing the MRC technique as

$$\hat{\mathbf{v}}_i = \hat{\mathbf{h}}_i^H \mathbf{Y'}_c. \tag{4.13}$$

Plugging (4.6) into (4.13), $\hat{\mathbf{v}}_i$ is written as

$$\hat{\mathbf{v}}_i = \hat{\mathbf{h}}_i^H \mathbf{h}_i \mathbf{v}_i + \mathbf{n}_i, \tag{4.14}$$

where $\mathbf{n}_i = \sum_{k \in \tilde{\mathcal{S}}_s, k \neq i} \hat{\mathbf{h}}_i^H \mathbf{h}_k \mathbf{v}_k + \hat{\mathbf{h}}_i^H \mathbf{Z}_c$. The first and second terms on the right-hand side of (4.14) are the signal and interference-plus-noise terms, respectively. We can approximate $\mathbf{n}_i$ to be Gaussian distributed, i.e., $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \sigma_{oi}^2 \mathbf{I}_{n_c})$, where $\sigma_{oi}^2 = \frac{1}{n_c}\mathbb{E}\{\mathbf{n}_i \mathbf{n}_i^H\} = P_c \sum_{k \in \hat{\mathcal{D}}_j, k \neq i} |\hat{\mathbf{h}}_i^H \mathbf{h}_k|^2 + \sigma_z^2 \|\hat{\mathbf{h}}_i\|^2$, which is obtained by treating the coded data sequences of different users to be uncorrelated. The demodulated signal can be obtained as

$$\mathbf{g}_i = \left[\mathrm{Im}\left(\vartheta_{1i}\right), \mathrm{Re}\left(\vartheta_{1i}\right), \ldots, \mathrm{Im}\left(\vartheta_{n_c i}\right), \mathrm{Re}\left(\vartheta_{n_c i}\right)\right], \tag{4.15}$$

where $\vartheta_{ti} = [\hat{\mathbf{v}}_i]_{(:,t)}$. From (4.14) and (4.15), and using $\hat{\mathbf{h}}_i^H \mathbf{h}_i \approx \|\hat{\mathbf{h}}_i\|^2$, each sample of $\mathbf{g}_i$ can be approximated as $\pm\sqrt{P_c/2}\|\hat{\mathbf{h}}_i\|^2 + n'$, where $n' \sim \mathcal{CN}\left(0, \frac{\sigma_{oi}^2}{2}\right)$. The following log-likelihood ratio (LLR) is obtained as the input to the polar list decoder

$$\mathbf{f}_i = \frac{2\sqrt{2P_c}\|\hat{\mathbf{h}}_i\|^2}{\hat{\sigma}_{oi}^2}\mathbf{g}_i, \tag{4.16}$$

where $\hat{\sigma}_{oi}^2$ is an approximation of $\sigma_{oi}^2$ which is obtained by replacing $\mathbf{h}_k$'s by their estimates. At the $j$th pilot part, the $i$th user is declared as successfully decoded if 1) its decoder output satisfies the CRC check, and 2) by mapping the $j$th pilot part of its decoded message to the Hadamard codebook, $\tilde{\mathbf{b}}_{jk}$ is obtained. Then, all the successfully decoded messages (in the current and previous iterations) are accumulated in the set $\mathcal{S}_s$, where $|\mathcal{S}_s| + |\tilde{\mathcal{S}}_s| = K_s$.

#### 4.3.2.4 SIC

we can see in (4.3) that the successfully decoded messages contain bit sequences of pilot parts and the coded part ($\mathbf{w}_{p_j}(i), j = 1, 2, ..., J$ and $\mathbf{w}_c(i)$). Having the bit sequences of successfully decoded messages, we can construct the corresponding transmitted signals using (4.4). The received signal matrix can be written as

$$\mathbf{Y} = \mathbf{H}_{\mathcal{S}_s}\mathbf{X}_{\mathcal{S}_s} + \mathbf{H}_{\tilde{\mathcal{S}}_s}\mathbf{X}_{\tilde{\mathcal{S}}_s} + \mathbf{Z}_s, \tag{4.17}$$

where $\mathbf{Y}$ is obtained by merging received signal matrices of different parts, i.e.,

$$\mathbf{Y} = [\mathbf{Y}_{p_1}, \ldots, \mathbf{Y}_{p_J}, \mathbf{Y}_c] \in \mathbb{C}^{M \times L}$$

with $\mathbf{X}_{\mathcal{S}_s} \in \mathbb{C}^{|\mathcal{S}_s| \times L}$ and $\mathbf{X}_{\tilde{\mathcal{S}}_s} \in \mathbb{C}^{|\tilde{\mathcal{S}}_s| \times L}$ including the signals in the sets $\mathcal{S}_s$ and $\tilde{\mathcal{S}}_s$, and $\mathbf{H}_{\mathcal{S}_s} \in \mathbb{C}^{M \times |\mathcal{S}_s|}$ and $\mathbf{H}_{\tilde{\mathcal{S}}_s} \in \mathbb{C}^{M \times |\tilde{\mathcal{S}}_s|}$ comprising the channel coefficients corresponding to the users in the sets $\mathcal{S}_s$ and $\tilde{\mathcal{S}}_s$, respectively. Employing the least squares (LS) technique, $\mathbf{H}_{\mathcal{S}_s}$ is estimated as

$$\hat{\mathbf{H}}_{\mathcal{S}_s} = \mathbf{Y}\mathbf{X}_{\mathcal{S}_s}^H(\mathbf{X}_{\mathcal{S}_s}\mathbf{X}_{\mathcal{S}_s}^H)^{-1}. \tag{4.18}$$

Note that $\mathbf{X}_{\mathcal{S}_s}$ consists of all the successfully decoded signals in the $s$th slot so far, and $\mathbf{Y}$ is the initially received signal matrix (not the output of the latest SIC

iteration). The SIC procedure is performed as follows

$$\mathbf{Y}' = [\mathbf{Y}'_{p_1}, \mathbf{Y}'_{p_2}, \ldots, \mathbf{Y}'_{p_J}, \mathbf{Y}'_c] = \mathbf{Y} - \hat{\mathbf{H}}_{\mathcal{S}_s}\mathbf{X}_{\mathcal{S}_s}. \tag{4.19}$$

Finally, $\mathbf{Y}'$ is fed back to the pilot detection algorithm for the next iteration, where the next pilot part is employed. We note that if no user is successfully decoded in $J$ consecutive iterations (corresponding to $J$ different pilot parts), the algorithm is stopped. The details of the decoding stages of MS-MRA are shown in Figure 4.2 and Algorithm 2. Note that we will discuss MS-MRA-WOPBE, which deviates from the above model, in Section 4.3.4.

**Theorem 1.** *The SINR at the output of MRC for a non-colliding user in the sth slot can be approximated as*

$$\beta_s \approx \frac{\omega_{c_s} P_c \left( \omega_{p_s} \mathbb{E}\{\|\mathbf{h}_i\|^4\} + \frac{\sigma_z^2}{n_p P_p} \mathbb{E}\{\|\mathbf{h}_i\|^2\} \right)}{\left( P_c(|\tilde{\mathcal{S}}_s| - 1) + \sigma_z^2 \right) \left( \omega_{p_s} \mathbb{E}\{\|\mathbf{h}_i\|^2\} + \frac{M\sigma_z^2}{n_p P_p} \right)}, \tag{4.20}$$

*where* $\omega_{p_s} = \omega_{c_s} = 1 - \frac{|\mathcal{S}_s|}{L}$ *if the transmitted signals are randomly interleaved, and* $\omega_{p_s} = 1 - \frac{1}{E_x} P_p |\mathcal{S}_s|$, $\omega_{c_s} = 1 - \frac{1}{E_x} P_c |\mathcal{S}_s|$, *otherwise, with* $E_x = J n_p P_p + n_c P_c$.

*Proof.*

**Lemma 2.** *Assuming that the transmitted data part contains uncorrelated and equally likely QPSK symbols, for* $i, j \in \mathcal{S}_s$ *and* $n_p, n_c \to \infty$, *the transmitted signals satisfy*

$$\frac{1}{E_x} \mathbf{x}_i \mathbf{x}_j^H \xrightarrow{p} 0, \tag{4.21}$$

*where* $E_x = J n_p P_p + n_c P_c$.

*Proof.* Let $\mathbf{b}_{ji}$ and $\mathbf{b}_{jr}$ be the $j$th pilots of the $i$th and $r$th users, and $\mathbf{v}_i$ and $\mathbf{v}_r$ be the corresponding polar-coded and QPSK-modulated signals. Since $\mathbf{b}_{ji}$ and $\mathbf{b}_{jr}$ are randomly chosen rows of the Hadamard matrix, $\mathbf{b}_{ji}\mathbf{b}_{ji}^T = n_p$ with probability

$\frac{1}{n_p}$, and it is zero with probability $1-\frac{1}{n_p}$. Besides, for $n_c \to \infty$, $v_{it}$ and $v_{rt}$ are zero-mean and uncorrelated, where $v_{it} = [\mathbf{v}_i]_{(:,t)}$. Therefore,

$$\begin{aligned}
&\lim_{n_p,n_c\to\infty} \mathbb{P}\left(\frac{1}{E_x}|\mathbf{x}_r\mathbf{x}_i^H|> 0\right) \\
&= \lim_{n_p,n_c\to\infty} \mathbb{P}\left(\frac{1}{E_x}\left|P_p\sum_{j=1}^{J}\mathbf{b}_{jr}\mathbf{b}_{ji}^H + \mathbf{v}_r\mathbf{v}_i^H\right| > 0\right) \\
&\leq \lim_{n_p,n_c\to\infty} \mathbb{P}\left(\frac{P_p}{E_x}\sum_{j=1}^{J}\left|\mathbf{b}_{jr}\mathbf{b}_{ji}^H\right| + \frac{1}{E_x}\left|\mathbf{v}_r\mathbf{v}_i^H\right| > 0\right) \\
&\leq \lim_{n_p,n_c\to\infty} \sum_{j=1}^{J}\mathbb{P}\left(\frac{P_p}{E_x}\left|\mathbf{b}_{jr}\mathbf{b}_{ji}^H\right| > 0\right) \\
&\qquad\qquad + \mathbb{P}\left(\frac{1}{E_x}\left|\sum_{t=1}^{n_c} v_{rt}v_{it}^H\right| > 0\right) \\
&\approx \lim_{n_p,n_c\to\infty} \frac{JP_p}{n_pE_x} + \mathbb{P}\left(\frac{n_c}{E_x}\left|\mathbb{E}\left\{v_{rt}v_{it}^H\right\}\right| > 0\right) \\
&\approx 0.
\end{aligned}$$

Note that, strictly speaking, the uncorrelated QPSK symbol assumption is not accurate for coded systems. Nevertheless, it is useful to obtain a good approximation of SINR, as we will show later. □

**Lemma 3.** *By applying LS-based SIC, the residual received signal matrices of pilot and coded parts can be written based on the signal and interference-plus-noise terms as*

$$\mathbf{Y}'_{p_j} \approx \sqrt{P_p}\mathbf{h}_i\mathbf{b}_{ji}\mathbf{L}_{p_j} + \sqrt{P_p}\sum_{k\in\tilde{\mathcal{S}}_s,k\neq i}\mathbf{h}_k\mathbf{b}_{jk}\mathbf{L}_{p_j} + \mathbf{Z}_{n,p_j}, \tag{4.22}$$

$$\mathbf{Y}'_c \approx \mathbf{h}_i\mathbf{v}_i\mathbf{L}_c + \sum_{k\in\tilde{\mathcal{S}}_s,k\neq i}\mathbf{h}_k\mathbf{v}_k\mathbf{L}_c + \mathbf{Z}_{n,c}, \tag{4.23}$$

*where $\mathbf{h}_i \in \mathbb{C}^{M\times 1}$ is the channel coefficient vector of the ith user, $\mathbf{L}_{p_j} = \omega_{p_s}\mathbf{I}_{n_p}$, $\mathbf{L}_c = \omega_{c_s}\mathbf{I}_{n_c}$, and the elements of $\mathbf{Z}_{n,p_j}$ and $\mathbf{Z}_{n,c}$ are drawn from $\mathcal{CN}(0,\omega_{c_s}\sigma_z^2)$ and $\mathcal{CN}(0,\omega_{p_s}\sigma_z^2)$, respectively, with $\omega_{p_s}$ and $\omega_{c_s}$ are as defined in the statement of the Theorem 1.*

*Proof.* Plugging (4.17) and (4.18) into (4.19), we obtain

$$\begin{aligned}\mathbf{Y}' &= \mathbf{H}_{\mathcal{S}_s}\mathbf{X}_{\mathcal{S}_s}\mathbf{L} + \mathbf{H}_{\tilde{\mathcal{S}}_s}\mathbf{X}_{\tilde{\mathcal{S}}_s}\mathbf{L} + \mathbf{Z}_s\mathbf{L} \\ &= \mathbf{H}_{\tilde{\mathcal{S}}_s}\mathbf{X}_{\tilde{\mathcal{S}}_s}\mathbf{L} + \mathbf{Z}_s\mathbf{L} \\ &= \mathbf{h}_i\mathbf{x}_i\mathbf{L} + \sum_{k\in\tilde{\mathcal{S}}_s, k\neq i}\mathbf{h}_k\mathbf{x}_k\mathbf{L} + \mathbf{Z}_n, \end{aligned} \tag{4.24}$$

where $\mathbf{L} = \mathbf{I}_L - \mathbf{X}_{\mathcal{S}_s}^H(\mathbf{X}_{\mathcal{S}_s}\mathbf{X}_{\mathcal{S}_s}^H)^{-1}\mathbf{X}_{\mathcal{S}_s}$, and $\mathbf{Z}_n = \mathbf{Z}_s\mathbf{L}$. Since $\mathbf{L}^H\mathbf{L} = \mathbf{L}$ and $\mathbf{Z}_s \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2\mathbf{I}_L)$, we have

$$\mathbf{Z}_n \sim \mathcal{CN}\left(\mathbf{0}, \sigma_z^2\mathbb{E}\{\mathbf{L}\}\right). \tag{4.25}$$

Since the values of $n_p$ and $n_c$ are large, and using (4.21), we have $\frac{1}{E_x}\mathbf{X}_{\mathcal{S}_s}\mathbf{X}_{\mathcal{S}_s}^H \approx \mathbf{I}_{|\mathcal{S}_s|}$, where $E_x = Jn_pP_p + n_cP_c$. In other words, we can approximate $\mathbf{L}$ as

$$\mathbf{L} \approx \mathbf{I}_L - \frac{1}{E_x}\sum_{r\in\mathcal{S}_s}\mathbf{x}_r^H\mathbf{x}_r. \tag{4.26}$$

Using the weak law of large numbers, and assuming samples of $\mathbf{x}_r$ to be uncorrelated and $|\mathcal{S}_l| \gg 1$, we can rewrite $\mathbf{L}$ in (4.26) as

$$\mathbf{L} \approx \begin{bmatrix} \mathbf{L}_{p_1} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \dots & \mathbf{L}_{p_J} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{L}_c \end{bmatrix}, \tag{4.27}$$

where $\mathbf{L}_{p_j} = \omega_{p_s}\mathbf{I}_{n_p}$ and $\mathbf{L}_c = \omega_{c_s}\mathbf{I}_{n_c}$ with $\omega_{p_s} = \omega_{c_s} = 1 - \frac{|\mathcal{S}_s|}{L}$ if the transmitted signals are randomly interleaved, and $\omega_{p_s} = 1 - \frac{1}{E_x}P_p|\mathcal{S}_s|$, $\omega_{c_s} = 1 - \frac{1}{E_x}P_c|\mathcal{S}_s|$, otherwise.

Letting $\mathbf{Z}_n = [\mathbf{Z}_{n,p_1}, \dots, \mathbf{Z}_{n,p_J}, \mathbf{Z}_{n,c}]$, we can infer from (4.25) and (4.27) that the elements of $\mathbf{Z}_{n,p_j}$ and $\mathbf{Z}_{n,c}$ approximately follow $\mathcal{CN}(0, \omega_{p_s}\sigma_z^2)$ and $\mathcal{CN}(0, \omega_{c_s}\sigma_z^2)$, respectively. Besides, using (4.27) and the signal structure in (4.4), we can divide (4.24) into pilot and coded parts as in (4.22) and (4.23). □

**Lemma 4.** *The estimated channel coefficients of a non-colliding user approximately satisfy the following expressions:*

$$\mathbb{E}\{\|\hat{\mathbf{h}}_i\|^2\} \approx \omega_{p_s}^2 \mathbb{E}\{\|\mathbf{h}_i\|^2\} + \frac{M\omega_{p_s}\sigma_z^2}{n_p P_p}, \tag{4.28a}$$

$$\mathbb{E}\{|\hat{\mathbf{h}}_i^H \mathbf{h}_k|^2\} \approx \omega_{p_s}^2 \mathbb{E}\{\|\mathbf{h}_i\|^2\} + \frac{M\omega_{p_s}\sigma_z^2}{n_p P_p}, \tag{4.28b}$$

$$\mathbb{E}\{|\hat{\mathbf{h}}_i^H \mathbf{h}_i|^2\} \approx \omega_{p_s}^2 \mathbb{E}\{\|\mathbf{h}_i\|^4\} + \frac{\omega_{p_s}\sigma_z^2}{n_p P_p} \mathbb{E}\{\|\mathbf{h}_i\|^2\}. \tag{4.28c}$$

*Proof.* Using the approximation of $\mathbf{Y}'_{p_j}$ in (4.22) in (4.12), the channel coefficient vector of the $i$th user can be estimated as

$$\begin{aligned}\hat{\mathbf{h}}_i &\approx \frac{\omega_{p_s}}{n_p}\mathbf{h}_i \mathbf{b}_{ji}\tilde{\mathbf{b}}_{jk}^H + \frac{\omega_{p_s}}{n_p}\sum_{f\in\tilde{\mathcal{S}}_s, f\neq i}\mathbf{h}_f \mathbf{b}_{jf}\tilde{\mathbf{b}}_{jk}^H + \mathbf{z}_{p_j,n} \\ &\overset{(a)}{=} \omega_{p_s}\mathbf{h}_i + \mathbf{z}_{p_j,n},\end{aligned} \tag{4.29}$$

where $\mathbf{z}_{p_j,n} = \frac{1}{n_p\sqrt{P_p}}\mathbf{Z}_{n,p_j}\tilde{\mathbf{b}}_{jk}^H$, and in (a), we use the assumption that the $i$th user is non-colliding, hence $\tilde{\mathbf{b}}_{jk}$ is only selected by the $i$th user ($\mathbf{b}_{ji} = \tilde{\mathbf{b}}_{jk}$ and $\mathbf{b}_{jf} \neq \tilde{\mathbf{b}}_{jk}^H$ for $f \in \tilde{\mathcal{S}}_s, f \neq i$). We can argue the following approximation $\mathbf{z}_{p_j,n} \sim \mathcal{CN}\left(0, \frac{\omega_{p_s}\sigma_z^2}{n_p P_p}\right)$. Using (4.29), we can show that $\mathbb{E}\{\|\hat{\mathbf{h}}_i\|^2\} \approx \omega_{p_s}^2\mathbb{E}\{\|\mathbf{h}_i\|^2\} + \frac{M\omega_{p_s}\sigma_z^2}{n_p P_p}$, $\mathbb{E}\{|\hat{\mathbf{h}}_i^H\mathbf{h}_i|^2\} \approx \omega_{p_s}^2\mathbb{E}\{\|\mathbf{h}_i\|^4\} + \frac{\omega_{p_s}\sigma_z^2}{n_p P_p}\mathbb{E}\{\|\mathbf{h}_i\|^2\}$, and $\mathbb{E}\{|\hat{\mathbf{h}}_i^H\mathbf{h}_k|^2\} = \mathbb{E}\{\|\hat{\mathbf{h}}_i\|^2\}$. □

Plugging (4.23) into the MRC expression in (4.13), $\hat{\mathbf{v}}_i$ can be estimated as

$$\hat{\mathbf{v}}_i \approx \omega_{c_s}\hat{\mathbf{h}}_i^H\mathbf{h}_i\mathbf{v}_i + \mathbf{z}_{in}, \tag{4.30}$$

where the first term on the right-hand side is the signal term, and $\mathbf{z}_{in} = \sum_{k\in\tilde{\mathcal{S}}_s, k\neq i}\hat{\mathbf{h}}_i^H\mathbf{h}_k\mathbf{v}_k\mathbf{L}_c + \hat{\mathbf{h}}_i^H\mathbf{Z}_{n,c}$ is the interference-plus-noise term. Since $\mathbf{L}^H\mathbf{L} = \mathbf{L}$, and using (4.27), we can show $\mathbf{L}_c^H\mathbf{L}_c \approx \mathbf{L}_c$. Therefore, by employing Lemma 4, we can approximate $\mathbf{z}_{in} \sim \mathcal{CN}(\mathbf{0}, \sigma_{in}^2\mathbf{I}_{n_c})$, where

$$\sigma_{in}^2 = \omega_{c_s}\left(P_c(|\tilde{\mathcal{S}}_s|-1) + \sigma_z^2\right)\left(\omega_{p_s}^2\mathbb{E}\{\|\mathbf{h}_i\|^2\} + \frac{M\omega_{p_s}\sigma_z^2}{n_p P_p}\right).$$

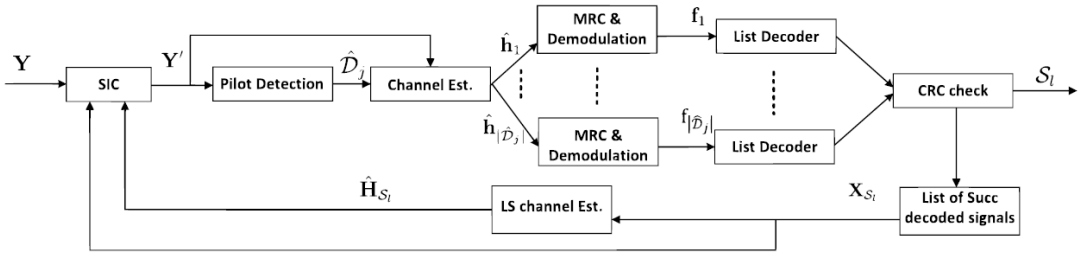


Figure 4.2: The decoding process of MS-MRA at the $j$th pilot part and the $s$th slot.

Besides, the per-symbol power of the signal term can be obtained as $\sigma_s^2 \approx \omega_{c_s}^2 \mathbb{E}\{|\hat{\mathbf{h}}_i^H \mathbf{h}_i|^2\} P_c$. Then, using Lemma 4, the SINR of $\hat{\mathbf{v}}_i$ can be calculated as in (4.20). □

We employ the above approximate SINR expression 1) to estimate the error probability of MS-MRA analytically, and 2) to determine the optimal power allocation for each group in MSUG-MRA. We further note that using this SINR approximation, the performance of the MS-MRA is well predicted in the low and medium $K_a$ regimes (see Figure 4.6). The reason why the SINR approximation does not work well in the high $K_a$ regime is the employed approximations in Lemma 3.

### 4.3.3 Analysis of MS-MRA

In this part, the PUPE of the MS-MRA is analytically calculated, where errors resulting from the collision, pilot detection, and polar decoder are considered. For our analyses, we assume that after successfully decoding and removing a user using a pilot part, the decoder moves to the next pilot part. Hence, in the $t$th iteration of the $s$th slot, we have

$$|\mathcal{S}_s| = t - 1, \tag{4.31a}$$

$$|\tilde{\mathcal{S}}_s| = K_s - t + 1. \tag{4.31b}$$

**Lemma 5.** *Let $\xi_k$ be the event that $k$ out of $K_s$ users remain in the $s$th slot, and define $\eta_i = \|\mathbf{h}_i\|^2$, where $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$. Assuming that the strongest users with*

**Algorithm 2:** The proposed MS-MRA decoder.

```
for l = 0, 1, ..., S do  Different slots
    S_s = ∅.
    flag = 1.
    t' = 0 (t' denotes the iteration index).
    while flag = 1 do
        t' = t' + 1.
        for j = 1, 2, ..., J do different pilot parts
            Pilot detection: estimate D̂_j using (4.8).
            Ch. estimation: estimate channel coefficient using (4.12).
            if MS-MRA then
                for i ∈ D̂_j do different detected pilots
                    MRC estimation: obtain v̂_i using (4.13).
                    Demodulation: obtain g_i using (4.15).
                    Decoding: pass f_i in (4.16) to list decoder.
                end
            end
            if MS-MRA-WOPBE then
                Perform IISD in Section 4.3.4.2.
            end
            S_{t'j}: set of successfully decoded users in the current iteration.
            S_s = S_s ⋃ S_{t'j}.
            LS-based ch. estimation: estimate Ĥ_{S_s} using (4.18).
            SIC: update Y'_{p_j} and Y'_c using (4.19).
        end
        if ⋃_{j=1}^{J} S_{t'j} = ∅ then
            flag = 0.
        end
    end
end
```

*highest $\eta_i$ values are decoded first, we have*

$$\mathbb{E}\{\eta_i^m|\xi_k\} = \mu_{(k,m)}, \tag{4.32}$$

*where* $\mu_{(k,m)} = \dfrac{\int_{-\infty}^{\bar{x}_k} \eta^m f_{2M}^{\chi^2}(2\eta)d\eta}{\int_{-\infty}^{\bar{x}_k} f_{2M}^{\chi^2}(2\eta)d\eta}$, *with* $f_k^{\chi^2}(.)$ *denoting the PDF of the chi-squared distribution with* $k$ *degrees of freedom and* $\bar{x}_k = 0.5\Gamma_{2M}^{-1}(k/K_s)$.

*Proof.* In the first iteration of the $s$th slot for which no user is decoded yet (all the $K_s$ active users are available), since $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, we have $2\eta_i|\xi_{K_s} \sim \chi_{2M}^2$. We assume that the users with higher values of $\eta_i$ are decoded first. Hence, if in an iteration, $k$ out of $K_s$ users remain in the slot, the distribution of $\eta_i$ is obtained by $2\eta_i|\xi_k \sim \{\chi_{2M}^2\}_{k/K_s}$, where $\{.\}_\beta$ removes $1-\beta$ portion of the samples with higher values from the distribution and normalizes the distribution of the remaining samples, i.e.,

$$\mathbb{P}(\eta_i = y|\xi_k) = \frac{f_{2M}^{\chi^2}(2y)}{\int_{-\infty}^{\bar{x}_k} f_{2M}^{\chi^2}(2y)dy}, y < \bar{x}_k,$$

where $\bar{x}_k$ is obtained by solving the following equation $\mathbb{P}\left(\eta_i < \bar{x}_k|\xi_{K_s}\right) = k/K_s$, which results in $\bar{x}_k = 0.5\Gamma_{2M}^{-1}(k/K_s)$. Therefore, we obtain

$$\mathbb{E}\{\eta_i^m|\xi_k\} = \frac{\int_{-\infty}^{\bar{x}_k} \eta^m f_{2M}^{\chi^2}(2\eta)d\eta}{\int_{-\infty}^{\bar{x}_k} f_{2M}^{\chi^2}(2\eta)d\eta}. \tag{4.33}$$

□

We can see from (4.15) and (4.16) that the input of the polar decoder is a $1 \times 2n_c$ real codeword. Thus, the average decoding error probability of a non-colliding user in the $t$th iteration of a slot with $K_s$ users can be approximated as (see [90])

$$P_{K_s,t}^{dec} \approx Q\left(\frac{0.5\log\left(1+\alpha_{K_s,t}\right) - \dfrac{B+r}{2n_c}}{\sqrt{\dfrac{1}{2n_c}\dfrac{\alpha_{K_s,t}(\alpha_{K_s,t}+2)\log^2 e}{2(\alpha_{K_s,t}+1)^2}}}\right), \tag{4.34}$$

where $Q(.)$ denotes the standard $Q$-function, and $\alpha_{K_s,t}$ is the SINR of a non-colliding user in the $t$th iteration of a slot with $K_s$ users, which is calculated using Theorem 1, Lemma 5, and (4.31) as

$$\alpha_{K_s,t} \approx \frac{s_{c_t} P_c \left( s_{p_t} \mu_{(K_s-t+1,2)} + \frac{\sigma_z^2}{n_p P_p} \mu_{(K_s-t+1,1)} \right)}{(P_c(K_s - t) + \sigma_z^2) \left( s_{p_t} \mu_{(K_s-t+1,1)} + \frac{M\sigma_z^2}{n_p P_p} \right)}, \tag{4.35}$$

where $s_{p_t} = 1 - P_p \frac{t-1}{E_x}$ and $s_{c_t} = 1 - P_c \frac{t-1}{E_x}$. Note that since the powers of signal and interference-plus-noise terms of $\hat{\mathbf{v}}_i$ are equal in their real and imaginary parts, the SINRs of $\mathbf{f}_i$ in (4.16) and $\hat{\mathbf{v}}_i$ are the same. Therefore, in (4.34), we employ the SINR calculated in Theorem 1 for the input of the polar list decoder.

Since decoding in the initial iterations well represents the overall decoding performance of the MS-MRA, we approximate the SINR of the first iteration by setting $t = 1$ in (4.35) as

$$\alpha_{K_s,1} \approx \frac{P_c M}{(\sigma_z^2 + P_c K_s) \left( 1 + \frac{\sigma_z^2}{n_p P_p} \right)}. \tag{4.36}$$

Concentrating on (4.34), we notice that $P_{K_s,1}^{dec}$ is a decreasing function of $n_c$ and $\alpha_{K_s,1}$. Besides, (4.36) shows that $\alpha_{K_s,1}$ increases by decreasing $n_c$ and $J$ (considering $K_s \approx K_a(Jn_p + n_c)/n$), and increasing $M$, $P_c$, and $P_p$, however, it is not a strictly monotonic function of $n_p$. Since our goal is to achieve the lowest $P_{K_s,t}^{dec}$ by spending the minimum $E_b/N_0 = (n_c P_c + J n_p P_p)/B$, we can optimize the parameters $n_c$, $n_p$, $P_c$, and $P_p$.

**Theorem 2.** *In the $t$th iteration of the $s$th slot, the probability of collision for a remaining user $i \in \tilde{\mathcal{S}}_s$ can be approximated as*

$$P_{K_s,t}^{col} \approx 1 - \frac{N_1^{(t)}}{K_s - t + 1}, \tag{4.37}$$

*where $N_i^{(k)}$ denotes the average number of pilots that are in $i$-collision (selected by $i$ different users) in the $k$th iteration, which is calculated as*

$$N_i^{(k+1)} \approx N_i^{(k)} + \begin{cases} \kappa_k \left( (i+1) N_{i+1}^{(k)} - i N_i^{(k)} \right) & i \geq 2 \\ \kappa_k \left( 2N_2^{(k)} - N_1^{(k)} \right) - \frac{1}{J} & i = 1 \end{cases}, \tag{4.38}$$

*where* $\kappa_k = \dfrac{J-1}{J(K_s-k+1)}$*, and* $N_i^{(1)} \approx n_p f_p(i; K_s/n_p)$ *with* $f_p(i;a)$ *denoting the probability mass function (PMF) of the Poisson distribution with the parameter* $a$*.*

*Proof.* In the first iteration of the $s$th slot, since $K_s$ users have selected one out of $n_p$ pilots randomly, the number of users that select an arbitrary pilot approximately follows a Poisson distribution with the parameter $K_s/n_p$. In the $k$th iteration of the $s$th slot, let $T_{j,i}^{(k)}$ be the average number of $i$-collision pilots (pilots selected by $i$ different users) in the $j$th pilot part. We have

$$T_{j,i}^{(1)} \approx n_p f_p(i; K_s/n_p), \tag{4.39}$$

where $f_p(i;a)$ denotes the PMF of the Poisson distribution with the parameter $a$. The average number of i-collision users in the $k$th iteration of the $j$th pilot part is then calculated as $K_{j,i}^{(k)} \approx iT_{j,i}^{(k)}$. Supposing that in the $k$th iteration (using the assumption in (4.31)), the decoder employs the $j$th pilot part for channel estimation, the removed user is non-colliding (1-collision) in its $j$th pilot part (we assume that the decoder can only decode the non-colliding users), and it is in $i$-collision in its $j'$th ($j' \neq j$) pilot part with probability $p_{i,j'}^{(k)} = \dfrac{K_{j',i}^{(k)}}{K_s-k+1}$. Therefore, removing a user from the $j$th pilot part results in

- In the $j$th pilot part, we have $T_{j,1}^{(k+1)} = T_{j,1}^{(k)} - 1$, and $T_{j,i}^{(k+1)} = T_{j,i}^{(k)}$ for $i > 1$.
- In the $j'$th pilot part ($j' \neq j$), we have $T_{j',i}^{(k+1)} = T_{j',i}^{(k)} + p_{i+1,j'}^{(k)} - p_{i,j'}^{(k)}$.

The collision probability of the $j$th pilot part in the $t$th iteration is then obtained as $P_{col}(j,t) = 1 - \dfrac{T_{j,1}^{(t)}}{K_s-t+1}$. Finally, by approximating $T_{j,i}^{(t)}$ by its average over different pilot parts (i.e., $T_{j,i}^{(t)} \approx N_i^{(t)} = \dfrac{1}{J}\sum_{j=1}^{J} T_{j,i}^{(t)}$) in above equations, the results in Theorem 2 are obtained. Note that since all the pilot parts are equally likely in the first iteration, we have $N_i^{(1)} \approx T_{j,i}^{(1)} \approx n_p f_p(i; K_s/n_p), \forall j = 1, ..., J$. □

Note that to extend the result in Theorem 2 to an SIC-based system with only one pilot sequence (orthogonal or non-orthogonal), we only need to set $J = 1$ in the above expressions. From (4.37), the collision probability in the first iteration can be calculated as $P_{K_s,1}^{col} \approx 1 - e^{-K_s/n_p}$, which is a decreasing function of $n_p$. Since the overall decoding performance of the system depends dramatically on the collision probability in the first iteration, we can increase $n_p$, however, this results in additional overhead.

**Corollary 2.** *Assuming a relatively large CRC length (hence negligible $p_{fa}$), the PUPE of the MS-MRA with $S$ slots and $K_a$ active users can be approximated as*

$$P_e \approx 1 - \sum_{r=1}^{K_a} (1 - \epsilon_r) \binom{K_a - 1}{r - 1} \left(\frac{1}{S}\right)^{r-1} \left(1 - \frac{1}{S}\right)^{K_a - r}, \tag{4.40}$$

*where $\epsilon_r$ denotes the PUPE of a slot with $r$ users, which is obtained as*

$$\epsilon_r \approx \sum_{j=1}^{r} \frac{r - j + 1}{r} p_{j,r}, \tag{4.41}$$

*with $p_{j,r} = (e_{j,r})^{r-j+1} \prod_{f=1}^{j-1} \left(1 - (e_{f,r})^{r-f+1}\right)$, and*

$$e_{t,r} = 1 - P_D(\delta_{NP}) \left(1 - P_{r,t}^{dec}\right) \left(1 - P_{r,t}^{col}\right), \tag{4.42}$$

*where $P_{r,t}^{dec}$, $P_{r,t}^{col}$, and $P_D(\delta_{NP})$ are computed in (4.34), Theorem 2, and (4.11), respectively.*

Note that the result in Corollary 2 can also be used in any other slotted system with SIC by replacing appropriate $e_{j,r}$.

### 4.3.4 MS-MRA-WOPBE

As discussed in Section 4.3.1, in the MS-MRA scheme, the pilot bits are fed to the polar encoder along with the data and CRC bits. To improve the performance by decreasing the coding rate, the MS-MRA-WOPBE scheme passes only the data and CRC bits to the encoder. To detect the bit sequences of different parts of the message, it employs an extra iterative decoding block called iterative inter-symbol

decoder (IISD) (described in Section 4.3.4.2). At each step of IISD, it detects one part of a user's signal (polar or pilot part), appends the detected part to the current pilot (which was used for channel estimation in the previous step) to have an extended pilot, and re-estimates the channel coefficients accordingly. The encoding and decoding procedures of MS-MRA-WOPBE are described below.

#### 4.3.4.1 Encoder

The $i$th user encodes its bits using the following steps (the general construction is shown in Figure 4.1). Similar to the MS-MRA encoder in Section 4.3.1, $B$ information bits are divided into $J+1$ parts as in (4.3), and the transmitted signal is generated as in (4.4). The only difference is in the construction of the QPSK signal. The encoder in MS-MRA-WOPBE defines two CRC bit sequences as $\mathbf{c}_2(i) = \mathbf{w}(i)\mathbf{G}_2$ and $\mathbf{c}_1(i) = [\mathbf{w}_c(i), \mathbf{c}_2(i)]\mathbf{G}_1$, where $\mathbf{G}_2 \in \{0,1\}^{B\times r_2}$ and $\mathbf{G}_1 \in \{0,1\}^{(B_c+r_2)\times r_1}$ are generator matrices known by the BS and users. Then, it passes $[\mathbf{w}_c(i), \mathbf{c}_2(i), \mathbf{c}_1(i)]$ to an $(2n_c,\ B_c + r_1 + r_2)$ polar encoder, and modulates the output by QPSK to obtain $\mathbf{v}_i \in \{\sqrt{P_c/2}(\pm 1 \pm j)\}^{1\times n_c}$.

#### 4.3.4.2 Decoder

As shown in Algorithm 2, MS-MRA-WOPBE exploits the same decoding steps as the MS-MRA scheme, except for the IISD step. We can see in Algorithm 2 that the $j$th pilot of the $i$th user is detected before employing the IISD. Then, IISD must detect the data (polar) sequence and the $f$th pilot of the $i$th user, where $f = 1, ..., J, f \neq j$. In the following, IISD is described in detail.

**Step 1** [Detecting $\mathbf{w}_c(i) \forall i \in \hat{\mathcal{D}}_j$]: We first obtain $\mathbf{g}_i$ using (4.15), where $\hat{\mathbf{v}}_i = \hat{\mathbf{h}}_i^H \mathbf{R}_h^{-1} \mathbf{Y}'_c$, and $\mathbf{R}_h = \sigma_z^2 \mathbf{I}_M + P_c \sum_{l\in\hat{\mathcal{D}}_j} \hat{\mathbf{h}}_l \hat{\mathbf{h}}_l^H$. Then, we pass $\mathbf{f}_i = \dfrac{2\sqrt{2P_c}}{1 - P_c \hat{\mathbf{h}}_i^H \mathbf{R}_h^{-1} \hat{\mathbf{h}}_i} \mathbf{g}_i$ to the list decoder. A CRC check $\text{flag}_{\text{CRC1}}(i) \in \{0,1\}$ and an estimate of $[\mathbf{w}_c(i), \mathbf{c}_2(i), \mathbf{c}_1(i)]$ [1] are obtained by the polar list decoder.

[1] In the output of the polar list decoder, there is a list of possible messages. If more than one messages satisfy the CRC check ($\mathbf{c}_1(i) = [\mathbf{w}_c(i), \mathbf{c}_2(i)]\mathbf{G}_1$), the most likely of them is returned

**Step 2** [Updating $\hat{\mathbf{h}}_i$]: Since the $j$th pilot and polar codeword of the $i$th user are detected so far, we append them to construct a longer signal as $\mathbf{q}_i = [\mathbf{b}_{ji}, \mathbf{v}_i] \in \mathbb{C}^{1\times(n_p+n_c)}$. Then, we update $\hat{\mathbf{h}}_i$ by MMSE estimation as $\hat{\mathbf{h}}_i = \mathbf{Y'}_q\mathbf{R}_q^{-1}\mathbf{q}_i^H$, where $\mathbf{R}_q = \sigma_z^2\mathbf{I}_{(n_p+n_c)} + \sum_{l\in\hat{\mathcal{D}}_j}\mathbf{q}_l^H\mathbf{q}_l$, and $\mathbf{Y'}_q = [\mathbf{Y'}_{p_j}, \mathbf{Y'}_c]$.

**Step 3** [Detecting $\mathbf{w}_{p_f}(i)\forall i \in \hat{\mathcal{D}}_j, f \neq j$]: Assuming that the $t$th row of the Hadamard matrix is active in the $f$th pilot part ($f \neq j$), we estimate the corresponding channel coefficient as $\mathbf{s}_{ft} = \dfrac{1}{n_p\sqrt{P_p}}\mathbf{Y'}_{p_f}\tilde{\mathbf{b}}_{ft}^T$ (see (4.12)). To find the $f$th pilot sequence of the $i$th user, we find the pilot whose corresponding channel coefficient vector is most similar to $\hat{\mathbf{h}}_i$, i.e., we maximize the correlation between $\hat{\mathbf{h}}_i$ and $\mathbf{s}_{ft}$ as

$$\hat{t}_{fi} = \max_t \frac{|\hat{\mathbf{h}}_i^H\mathbf{s}_{ft}|^2}{\mathbf{s}_{ft}^H\mathbf{s}_{ft}}, f = 1, ..., J, f \neq j. \tag{4.43}$$

**Step4** [Updating $\hat{\mathbf{h}}_i$]: Since the bit sequences of all $J+1$ parts are detected, we can construct $\mathbf{x}_i$ using (4.4). The channel coefficient vector can be updated by MMSE as $\hat{\mathbf{h}}_i = \mathbf{Y'}\mathbf{R}^{-1}\mathbf{x}_i^H$, where $\mathbf{R} = \sigma_z^2\mathbf{I}_L + \sum_{l\in\hat{\mathcal{D}}_j}\mathbf{x}_l^H\mathbf{x}_l$. If the number of users that satisfy $\text{flag}_{\text{CRC1}}(i) = 1$ is not changed in an iteration, the iteration is stopped, otherwise, the algorithm goes to Step 1 for another iteration with updated $\hat{\mathbf{h}}_i$. Users whose bit sequences satisfy $\mathbf{c}_2(i) = \mathbf{w}(i)\mathbf{G}_2$ and $\mathbf{c}_1(i) = [\mathbf{w}_c(i), \mathbf{c}_2(i)]\mathbf{G}_1$ are added to the set $\mathcal{S}_{t'j}$ as successfully decoded users of the current iteration.

### 4.3.5 MSUG-MRA

Different from MS-MRA where the power of every user is the same and signals are not interleaved, MSUG-MRA defines $G$ groups, each being assigned unique interleaver and power pair $(\pi_g(.), P_{p_g}, P_{c_g})$, $g = 1, 2, ..., G$. We assume that $\phi = \dfrac{P_{p_g}}{P_{c_g}}$ is constant in all groups, hence each group can be identified with a unique interleaver-power pair $(\pi_g(.), P_{c_g})$, which is known at both transmitter and receiver sides. The details of encoding and decoding procedures as well as the power selection strategy are explained below. Note that we assume without loss of generality that $P_{c_1} < P_{c_2} ... < P_{c_G}$.

as the detected message and the CRC flag is set to one. Otherwise, the most likely message is returned as the detected message and the CRC flag is set to zero.

#### 4.3.5.1 Encoder

The encoding is adopted as follows:

- Every user randomly selects a group, e.g., with index $g$.
- Each user employs $P_{c_g}$ and $\phi P_{c_g}$ as the powers of the coded and pilot parts, with which it generates its multi-stage signal $\mathbf{x}_i$ similar to MS-MRA (according to (4.4)).
- The transmitted signal is created as $\tilde{\mathbf{x}}_i = \pi_g(\mathbf{x}_i)$.

#### 4.3.5.2 Decoder

In each iteration, the decoder tends to decode the messages belonging to the users of the dominant group (the $G$th group with the highest power level). After decoding and removing users in the $G$th group, users in the $(G-1)$st group become the dominant ones. Using the same trend, all the groups have the chance to be the dominant group at some point. Since users in different groups are interleaved differently, signals of users in other groups are uncorrelated from the signals in the dominant group. Thus, letting the $g_0$th group to be dominant, we approximately model the $f$th signal in the the $g$th group ($g \neq g_0$) as

$$\tilde{\mathbf{x}}_f \sim \mathcal{CN}(\mathbf{0}, \zeta P_{c_g} \mathbf{I}_L), \tag{4.44}$$

where $\zeta = \dfrac{J\phi n_p + n_c}{L}$. Therefore, when the $g_0$th group is dominant (the users in the groups with indices greater than $g_0$ are already removed using SIC), users in the $g_0$th group are perturbed by i.i.d. noise samples drawn from $\mathcal{CN}(0, \delta_{g_0})$, with $\delta_{g_0} \approx \zeta K_0 \sum_{g=1}^{g_0-1} P_{c_g} + \sigma_z^2$, where $K_0 = \dfrac{K_a}{SG}$ is the average number of users in each group of the current slot. Consequently, by replacing $\sigma_z^2$, $P_p$, and $P_c$ with $\delta_{g_0}$, $\phi P_{c_{g_0}}$, and $P_{c_{g_0}}$ in the decoding steps of MS-MRA (in Section 4.3.2), the decoding procedure of MSUG-MRA is obtained as:

- Deinterleave the rows of the received signals: $\tilde{\mathbf{Y}}'_{p_j} = \pi_{g_0}^{-1}(\mathbf{Y}'_{p_j})$ and $\tilde{\mathbf{Y}}'_c = \pi_{g_0}^{-1}(\mathbf{Y}'_c)$.

- Find active pilots as

$$\hat{\mathcal{D}}_j = \left\{ l : \tilde{\mathbf{u}}_{jl}^H \tilde{\mathbf{u}}_{jl} \geq 0.5\delta_{g_0}\Gamma_{2M}^{-1}(1-\gamma) \right\},$$

  where $\tilde{\mathbf{u}}_{ji} = \tilde{\mathbf{Y}}'_{p_j}\bar{\mathbf{b}}_i^H/\sqrt{n_p}$.

- Channel estimation and MRC: $\hat{\mathbf{v}}_i = \hat{\mathbf{h}}_i^H \tilde{\mathbf{Y}}'_c$, where $\hat{\mathbf{h}}_i = \dfrac{1}{n_p\sqrt{\phi P_{c_{g_0}}}}\tilde{\mathbf{Y}}'_{p_j}\tilde{\mathbf{b}}_{jk}^T$, and $\tilde{\mathbf{b}}_{jk}$ is one of the detected pilots.

- Pass $\mathbf{f}_i = \dfrac{2\sqrt{2P_{c_{g_0}}}\|\hat{\mathbf{h}}_i\|^2}{\hat{\sigma}_{oi}^2}\mathbf{g}_i$ to the polar decoder, where $\hat{\sigma}_{oi}^2 = P_{c_{g_0}}\sum_{k\in\hat{\mathcal{D}}_j, k\neq i}|\hat{\mathbf{h}}_i^H\hat{\mathbf{h}}_k|^2 + \delta_{g_0}\|\hat{\mathbf{h}}_i\|^2$, and $\mathbf{g}_i$ is defined in (4.15).

- Regenerate signals of successfully decoded users according to Section 4.3.5.1 (using $(\pi_{g_0}(.), P_{c_{g_0}})$ pair), and collect them in the rows of $\tilde{\mathbf{X}}_{\mathcal{S}_s}$.

- Apply LS-based SIC similar to (4.19), i.e., $\mathbf{Y}' = \mathbf{Y}(\mathbf{I}_L - \tilde{\mathbf{X}}_{\mathcal{S}_s}^H(\tilde{\mathbf{X}}_{\mathcal{S}_s}\tilde{\mathbf{X}}_{\mathcal{S}_s}^H)^{-1}\tilde{\mathbf{X}}_{\mathcal{S}_s})$.

Note that this loop is repeated for $G$ different group indices and $J$ different pilot parts, and the iteration is stopped if there is no successfully decoded users in $GJ$ consecutive iterations.

#### 4.3.5.3 Power Calculation

When MSUG-MRA starts the decoding in the $g_0$th group, there are $|\mathcal{S}_s| \approx K_0(G-g_0)$ successfully decoded users from previous groups (with higher power levels), $|\tilde{\mathcal{S}}_s| = K_0$ users remain in the $g_0$th group, and users in the current group are perturbed with a complex Gaussian noise with covariance matrix $\delta_{g_0}\mathbf{I}_M$. Therefore, the SINR of a non-colliding user in the current group can be calculated by replacing $|\tilde{\mathcal{S}}_s| \approx K_0$, $|\mathcal{S}_s| = K_0(G-g_0)$, $\mathbb{E}\{\|\mathbf{h}_i\|^2\} = M$, $\mathbb{E}\{\|\mathbf{h}_i\|^4\} = M^2$, $P_c = P_{c_{g_0}}$, $P_p = \phi P_{c_{g_0}}$, $\sigma_z^2 \approx \delta_{g_0}$, and $\omega_{p_s} = \omega_{c_s} = 1 - \dfrac{|\mathcal{S}_s|}{L}$ in (4.20) as

$$\beta'_{g_0} \approx \frac{\rho_{g_0} M P_{c_{g_0}}^2 + \dfrac{\delta_{g_0}}{n_p\phi}P_{c_{g_0}}}{\left(P_{c_{g_0}}(K_0-1)+\delta_{g_0}\right)\left(P_{c_{g_0}} + \dfrac{\delta_{g_0}}{\rho_{g_0} n_p \phi}\right)}, \tag{4.45}$$

where $\rho_{g_0} = 1 - \dfrac{K_0(G-g_0)}{L}$. To impose similar performance on different groups, we set $\beta'_1 = \beta'_2 = \ldots = \beta'_G$. Solving this equation, the power of the $g$th group satisfies $c_1 P_g^2 + c_2 P_g + c_3 = 0$, where $c_1 = (K_0 - 1) - \dfrac{\rho_g M}{\beta'_{g-1}}$, $c_2 = \delta_g \left(1 + \dfrac{(K_0-1)}{\phi n_p \rho_g} - \dfrac{1}{\phi n_p \beta'_{g-1}}\right)$, $c_3 = \dfrac{\delta_g^2}{\phi n_p \rho_g}$. Solving this equation, we have

$$P_t = \frac{-c_2 + \sqrt{c_2^2 - 4c_1c_3}}{2c_1}, \tag{4.46}$$
$$\text{s.t. } \frac{1}{G}\sum_{f=1}^{G} P_f = P \text{ and } \; P_t \in \mathbb{R}^+.$$

Note that the MS-MRA scheme is a special case of the MSUG-MRA with $G = 1$.

### 4.3.6 MS-SRA and MSUG-SRA

In this part, we apply the proposed MIMO coding schemes to the case of a single receive antenna. To accomplish this, we repeat each user's length-$L$ signal multiple times to create temporal diversity in MS-SRA and MSUG-SRA. Accordingly, we divide the whole frame into $V$ sub-frames of length $n' = n/V$, then divide each sub-frame into $S$ slots of length $L = n'/S$. Each user randomly selects a slot index, namely $s$, and transmits its signal, through the $s$th slot of each sub-frame. Assuming the coherence time to be $L$, each sub-frame is analogous to a receive antenna. Therefore, the transmitted messages in MS-SRA and MSUG-SRA can be decoded using MS-MRA and MSUG-MRA decoders in Sections 4.3.2 and 4.3.5.2, respectively, considering $V$ receive antennas. Since each user repeats its signal $V$ times, for this case, we have $E_b/N_0 = \dfrac{VLP}{\sigma_z^2 B}$.

### 4.3.7 Computational Complexity

We focus on the number of multiplications as a measure of the computational complexity, and make a complexity comparison among the proposed and existing URA solutions. The per-iteration computational complexity of the MS-MRA in

a slot is calculated as follows: The pilot detection in (4.8) has a complexity of $\mathcal{O}(n_p^2 MJS)$ corresponding to $J$ different pilot parts and $S$ different slots, where $\mathcal{O}(.)$ is the standard big-O notation, denoting the order of complexity. The channel estimator in (4.12) does not require any extra computation, because $\hat{\mathbf{h}}_i$ corresponds to $\mathbf{u}_{ji}$ which is calculated before for pilot detection; the MRC in (4.13) has a complexity of $\mathcal{O}(\sum_{j=1}^{J}|\mathcal{D}_j|Mn_cS)$; to compute the LLR in (4.16), the required computational complexity is $\mathcal{O}(\sum_{j=1}^{J}|\mathcal{D}_j|^2MS)$; the computational complexity of the polar list decoder is [91] $\mathcal{O}(\sum_{j=1}^{J}|\mathcal{D}_j|n_c\log n_cS)$; and, the SIC has a complexity of $\mathcal{O}(ML|\mathcal{S}_s|S+|\mathcal{S}_s|^2LS)$. We know from (4.39) that in the first iteration, we have $|\mathcal{D}_j|\approx n_p - n_pe^{-K_a/(n_pS)} < K_a/S$, and $|\mathcal{S}_s|= 0$; in the last iterations, we have $|\mathcal{S}_s|\approx K_a/S$ and $|\mathcal{D}_j|\approx 0$. Hence, considering $M \gg \log n_c$ and $n_c|\mathcal{D}_j|\gg n_p$, we can compute the computational complexity of the MS-MRA in the first and last iterations as $\mathcal{O}\left(K_aMJ(n_c+K_a/S)\right)$ and $\mathcal{O}\left(LK_a(M+K_a/S)\right)$, respectively. Considering the computational complexity in the intermediate iterations to be in the same order, the per-iteration computational complexity of the MS-MRA can be bounded by $\mathcal{O}\Big(n_p^2MJS+\max\left(K_aMJ(n_c+K_a/S), LK_a(M+K_a/S)\right)\Big)$. Note that the computational complexity of MSUG-MRA is in the same order as MS-MRA, and for MS-SRA and MSUG-SRA schemes, the computational complexity is obtained by replacing $M$ by $V$ in the above figures.

Looking at Algorithm 2, we can infer that MS-MRA-WOPBE is obtained by employing the same pilot detector (with complexity $\mathcal{O}(n_p^2MJS)$), channel estimator (does not incur any extra computational complexity), and SIC (with complexity $\mathcal{O}(ML|\mathcal{S}_s|S+|\mathcal{S}_s|^2LS)$) as in the MS-MRA case, except for employing the IISD block. In Step 1 of IISD, the complexity for computing $\mathbf{f}_i$ and implementing polar decoder are $\mathcal{O}((Mn_c+M^2)T_IS\sum_{j=1}^{J}|\mathcal{D}_j|)$ and $\mathcal{O}(T_In_c\log n_cS\sum_{j=1}^{J}|\mathcal{D}_j|)$, respectively, where $T_I$ denotes the number of iterations of IISD. In the Step 2 of IISD, computing $\hat{\mathbf{h}}_i$ and $e_k$ has the complexity of $\mathcal{O}(T_I(n_c+n_p)^2S\sum_{j=1}^{J}|\mathcal{D}_j|+T_I(n_c+n_p)MS\sum_{j=1}^{J}|\mathcal{D}_j|)$ and $\mathcal{O}(T_I(J-1)n_pMS\sum_{j=1}^{J}|\mathcal{D}_j|)$, respectively. The computational complexity of obtaining $\hat{\mathbf{h}}_i$ in Step 3 of IISD is $\mathcal{O}(T_I(L^2+LM)S\sum_{j=1}^{J}|\mathcal{D}_j|)$. Then, replacing $|\mathcal{D}_j|$ and $|\mathcal{S}_s|$ with their approximate values (discussed in the previous paragraph), the overall computational complexity of the MS-MRA-WOPBE is bounded by $\mathcal{O}\Big(n_p^2MJS+\max\left(((L^2+M^2)+ML)\,T_IJK_a, LK_a(M+K_a/S)\right)\Big)$.

For comparison purposes, the dominant per-iteration computational complexity of the FASURA in [73] (which is due to energy detector and SIC operation) can also be computed as $\mathcal{O}\left(M(n_p + L'n_c)2^{B_f} + K_a(nM + n^2)\right)$, where $B_f$ denotes the number of pilot bits, $n$ is the frame length, and $L'$ is the length of the spreading sequence.

## 4.4 Numerical Results

We provide a set of numerical results to assess the performance of the proposed URA set-ups. In all the results, we set $B = 100$, the number of CRC bits $r = 11$, the Neyman-Pearson threshold $\gamma = 0.1$, and the list size of the decoder to 64. For MS-MRA and MSUG-MRA, we set the frame length $n \approx 3200$, and $P_e = 0.05$. The corresponding values for the MS-SRA and MSUG-SRA are $n \approx 30000$, and $P_e = 0.1$.

In Figure 4.3, the performance of the proposed MS-MRA and MSUG-MRA is compared with the short blocklength scheme of [56] with the number of antennas $M = 100$ and slot length $L = 200$. (In this scenario, we consider a fast-fading environment, where the coherence blocklength is considered as $L_c = 200$). To facilitate a fair comparison, we consider $(J, n_p, n_c) = (2, 32, 128)$ ($L = 192$) and $P_p/P_c = 1$ ($\phi = 1$ for MSUG-MRA) for all the proposed schemes. For MSUG-MRA, the value of $G$ is set as $G = 1$ for $K_a \leq 400$, $G = 3$ for $K_a = 500$, $G = 6$ for $600 \leq K_a \leq 800$, $G = 8$ for $900 \leq K_a \leq 1000$, and $G = 10$ for $K_a > 1000$. The superiority of the proposed schemes over the one in [56] is mostly due to the more powerful performance of the polar code compared to the simple coding scheme adopted in [56] and the use of the SIC block, which significantly diminishes the effect of interference. We also observe that MS-MRA-WOPBE outperforms MS-MRA, which is due to 1) employing IISD, which iteratively improves the accuracy of the channel estimation, and 2) lower coding rate by not encoding the pilot bits. Besides, the range of the number of active users that are detected by the MSUG-MRA is higher than those of MS-MRA and MS-MRA-WOPBE schemes. This improvement results from randomly dividing users into different groups, which provides each group with a lower number of active users (hence a lower effective

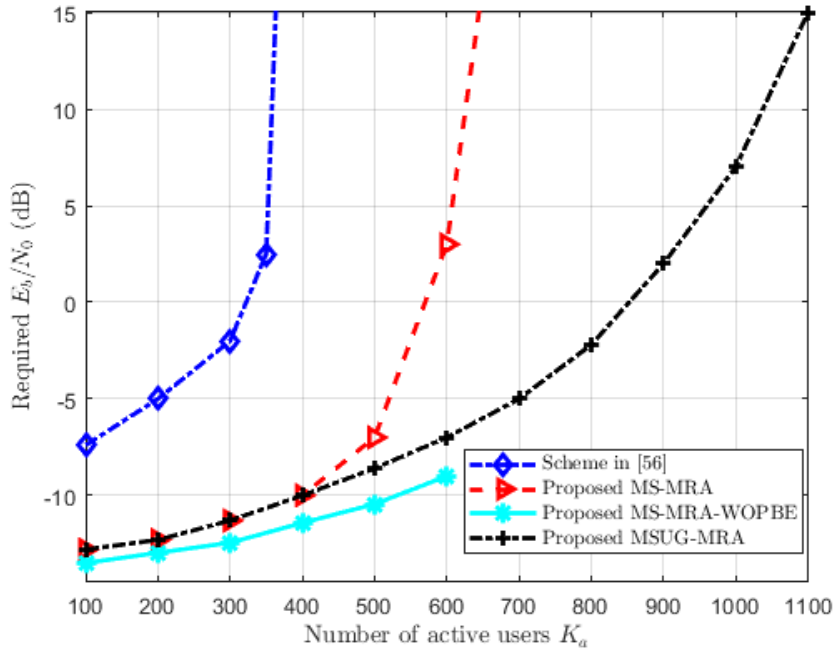


Figure 4.3: The required $E_b/N_0$ in the proposed MIMO set-ups and the scheme in [56] for $L \approx 200$, $M = 100$, and $P_e = 0.05$.

interference level).

In Figure 4.4, we compare the proposed MS-MRA and MSUG-MRA with the ones in [72, 73], considering the slow-fading channel with coherence blocklength $L_c = 3200$. We set $(J, n_p, n_c) = (2, 256, 512)$, $M = 50$, $P_p/P_c = 0.66$ for MS-MRA. We choose $(J, n_p, n_c, G) = (2, 256, 512, 1)$ for $K_a \leq 700$, $(J, n_p, n_c, G) = (2, 64, 512, 6)$ for $K_a = 900$, and $(J, n_p, n_c, G) = (2, 64, 512, 18)$ for $K_a > 900$ with $\phi = 0.66$. Thanks to employing the slotted structure, SIC, and orthogonal pilots, all the proposed schemes have superior performance compared to [72]. Due to employing random spreading and an efficient block called NOPICE, FASURA in [73] performs better than the proposed MS-MRA and MSUG-MRA in the low $K_a$ regimes; however, its performance is worse than the MSUG-MRA in higher values of $K_a$ (thanks to the random user grouping employed in MSUG-MRA). The proposed MS-MRA-WOPBE also shows a similar performance as FASURA. To achieve the result in Figure 4.4, FASURA sets $n_p = 896$, $L' = 9$, $n_c = 256$, $n = 3200$, $B_f = 16$, and $M = 50$. The order of computational complexity for

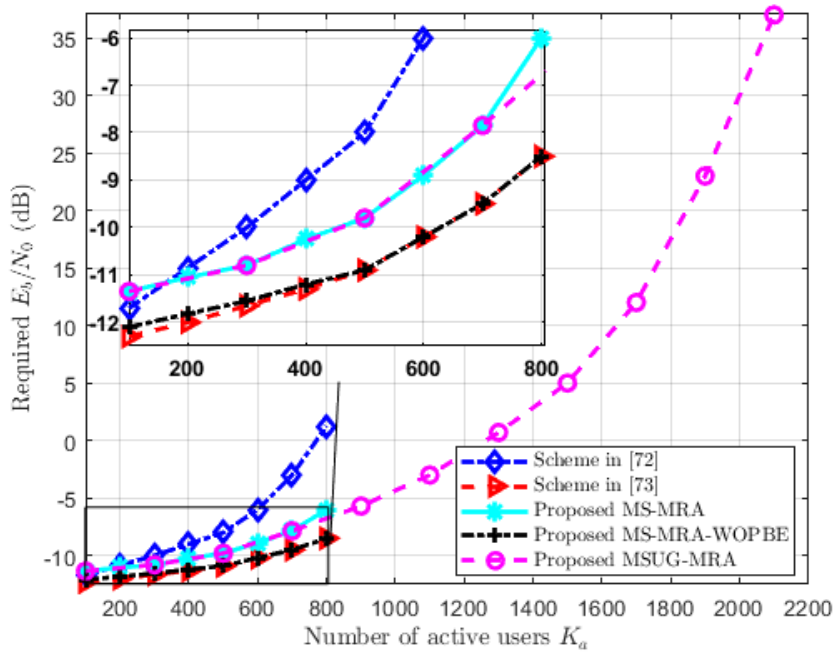


Figure 4.4: The required $E_b/N_0$ in the proposed MIMO set-ups and the results in [72, 73] for $M = 50$.

these schemes is given in the performance-complexity plot in Figure 4.5. It can be interpreted from this figure that the proposed MS-MRA-WOPBE has comparable accuracy to FASURA while offering a lower computational complexity. Note also that despite the higher required $E_b/N_0$ compared to FASURA, MS-MRA offers very large savings in terms of computational complexity, which is attributed to employing orthogonal pilots, slotted structure, and simpler decoding blocks.

As a further note, FASURA considers $2^{B_p}$ possible spreading sequences of length $L'$ for each symbol of the polar codeword; hence every transceiver should store $n_c 2^{B_p}$ vectors of length $L'$, as well as a pilot codebook of size $2^J \times n_p$. For typical values reported in [73], the BS and every user must store $1.6 \times 10^7$ vectors of length 9 and a matrix of size $5.8 \times 10^7$. For our proposed scheme, every transceiver must store only an orthogonal codebook of size $n_p \times n_p$, where $n_p = 256$. Thus, FASURA requires about 3000 times larger memory than our proposed schemes, which may be restrictive for some target URA applications such as sensor networks, where a massive number of cheap sensors are deployed. Moreover, unlike

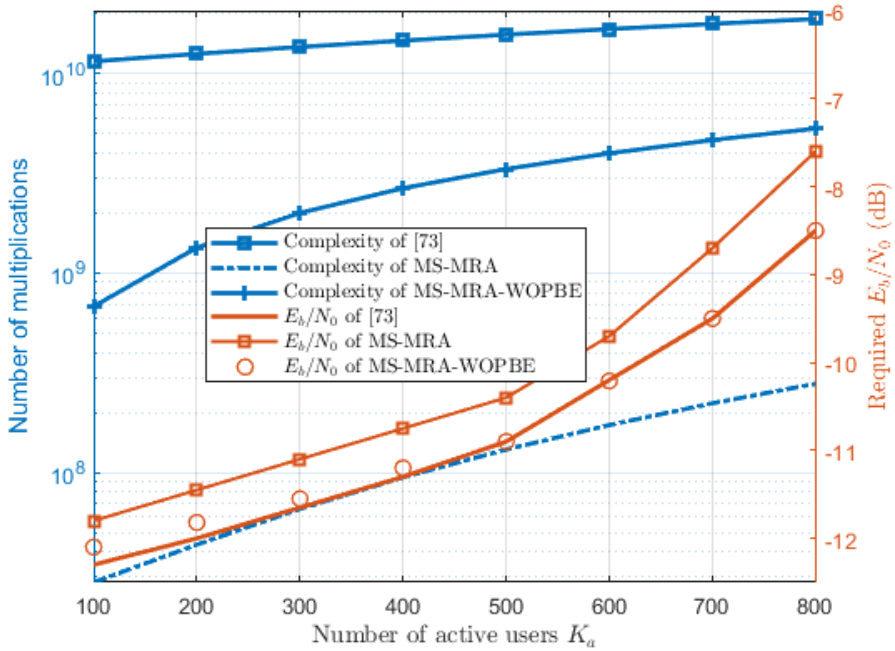


Figure 4.5: Performance-complexity curve for the proposed MIMO schemes and FASURA in [73].

FASURA, the proposed solutions are implementable with short blocklengths (see Figure 4.3), which makes them appropriate for fast fading scenarios as well.

In Figure 4.6, we compare the theoretical PUPE in (4.40) with the simulation results of the MS-MRA for three different scenarios ($M = 50, 100, 200$) with $P_p/P_c = 0.66$ and $(J, n_p, n_c) = (2, 256, 512)$. It is shown that the approximate theoretical analysis well predicts the performance of the MS-MRA for $K_a \leq 700$, however, the results are not consistent for higher values of $K_a$. The reason for the mismatch for the $K_a > 800$ regime is the approximations employed while analyzing SIC in Lemma 3 (e.g., $n_c, n_p \gg 1$, $|\mathcal{S}_s| \gg 1$, uncorrelated QPSK codewords of two different users, and uncorrelated samples of $\mathbf{x}_i$). Besides, in Figure 4.7 we compare the simulated and theoretical results of MS-MRA for different frame-length values. We can see in this figure that the prediction accuracy of our theoretical analysis well matches the simulation when $L$ is large (the region where the approximation in (4.27) is valid).

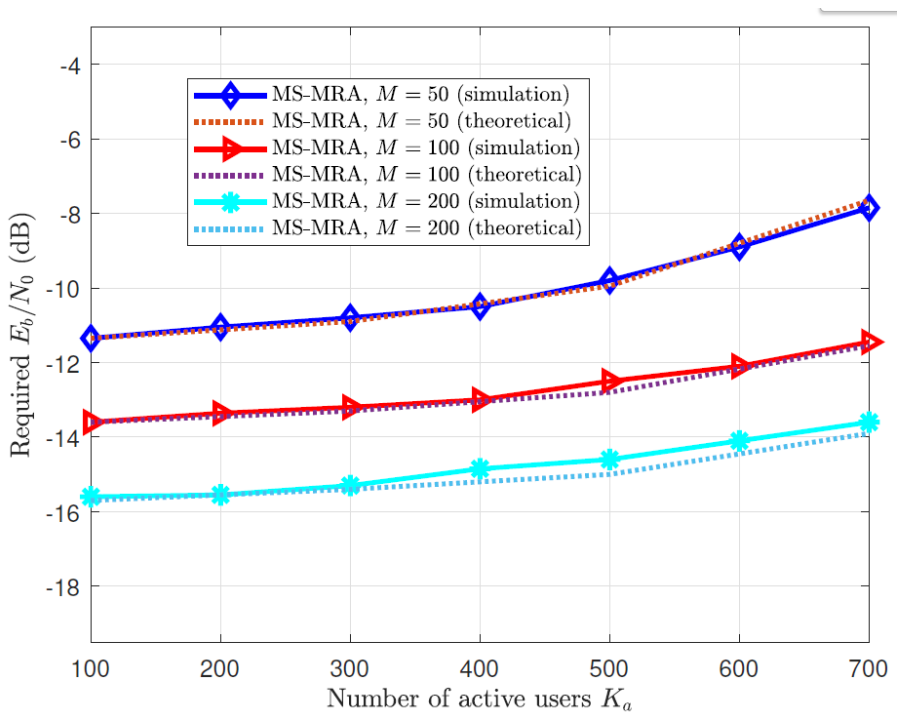


Figure 4.6: Comparison of the simulation and analytical performance of the MS-MRA for different values of $M$.

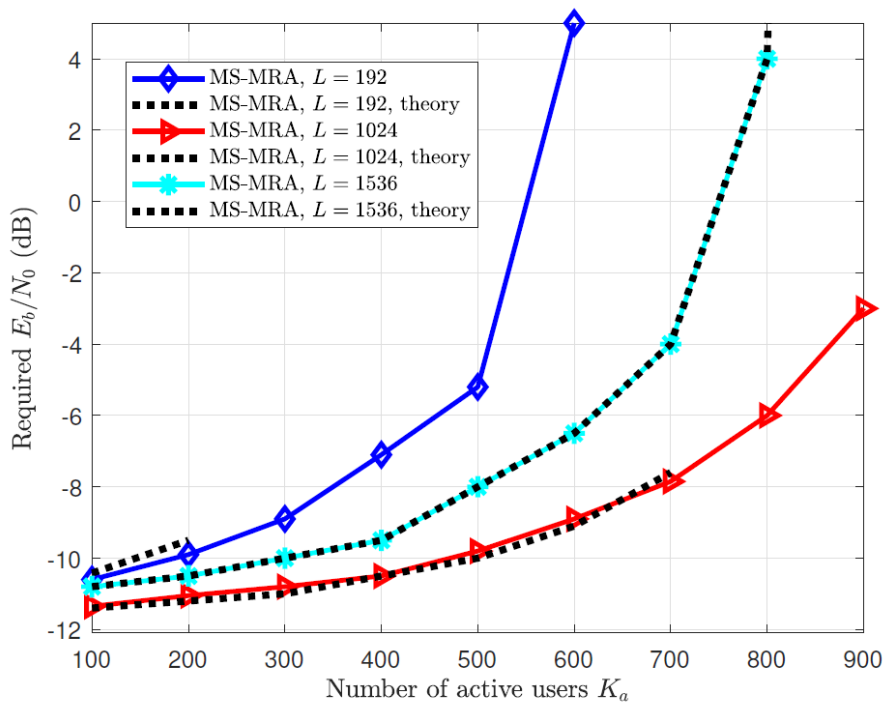


Figure 4.7: Comparison of the simulation and analytical performance of the MS-MRA for different values of blocklength.

Figure 4.8 compares the MS-SRA and MSUG-SRA with the existing single-antenna solutions [68, 80, 86]. For both set-ups, we set $(J, n_p, n_c) = (2, 64, 512)$, $P_p/P_c = 1$ ($\phi = 1$ for MSUG-SRA), $(S, V) = (6, 8)$ for $K_a \leq 200$, and $(S, V) = (12, 4)$ for $K_a \geq 300$. For MSUG-SRA, we also choose $G = 1$ for $K_a \leq 300$, $G = 3$ for $500 \leq K_a \leq 700$, and $G = 6$ for $K_a \geq 900$. It is observed that the proposed MS-SRA has a superior performance compared to the existing URA approaches for the low number of active users, However, it performs worse than the scheme in [80] for higher values of $K_a$. Furthermore, the proposed MSUG-SRA outperforms existing solutions, and its effective range of $K_a$ is up to 1500 users.

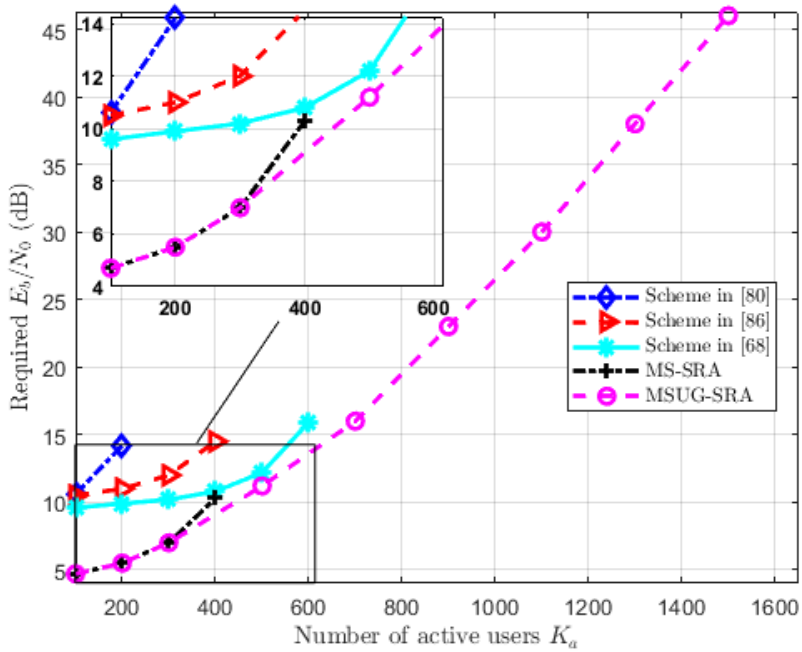


Figure 4.8: The required $E_b/N_0$ of the proposed MS-SRA and MSUG-SRA for the case of single-antenna receiver.

In order to validate our derivations in Theorem 2, Figure 4.9 compares $P^{col}_{K_s,t}$ in (4.37) and its corresponding values derived from Monte Carlo simulations. We conducted these simulations with specific parameters, namely, $JK_s = 300$, $Jn_p = 512$, and different values of $J$. The results indicate a strong alignment between the theoretical calculations and the empirical observations. Furthermore, it is noteworthy that increasing the value of $J$ leads to a more pronounced reduction in collisions.

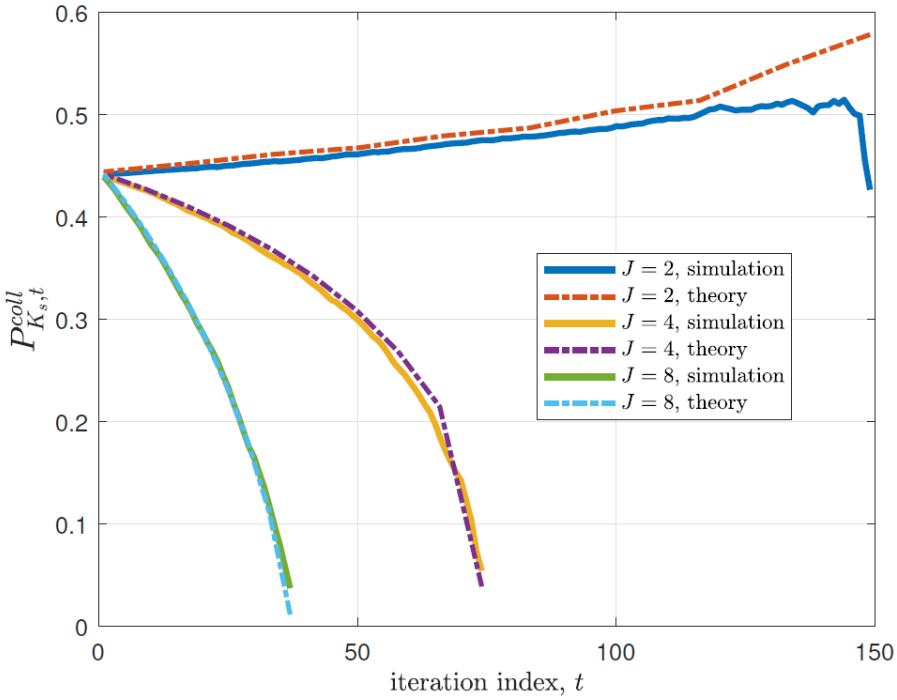


Figure 4.9: Comparison between the theoretical and simulated values of collision probability.

In Figure 4.10, we have plotted the required $E_b/N_0$ of the MS-MRA for different values of $J$ such that $Jn_p = 512$. Note that due to employing the Hadamard matrix in our pilot design, the pilot length needs to be a power of 2. We can see in this figure that for $K_a > 100$, $J = 2$ is the efficient selection, while for $K_a \leq 100$, $J = 1$ is the optimal choice. This figure clearly shows the trade-off in selecting $J$. Also, to show the effect of the parameter $n_p$ on the performance of the proposed scheme, we provide several examples for $n_c = 512, J = 2, M = 50$, $P_c/P_p = 1.5$ in Figure 4.11. It is observed that the performance of the decoder is highly sensitive to this parameter, especially, for larger values of $K_a$.

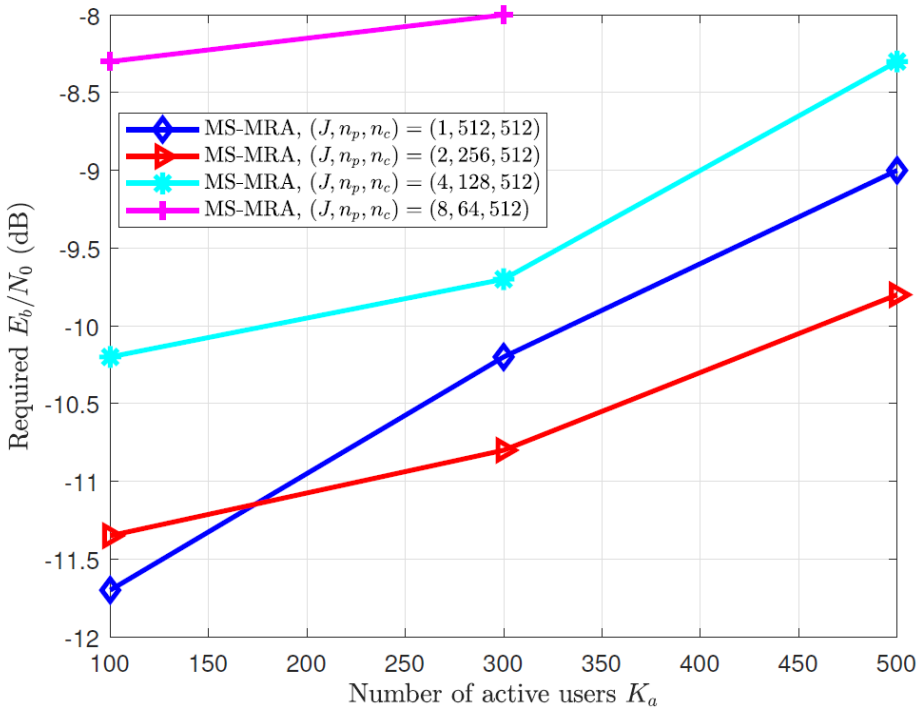


Figure 4.10: The required $E_b/N_0$ of MS-MRA for different $J$ values with fixed $Jn_p$.

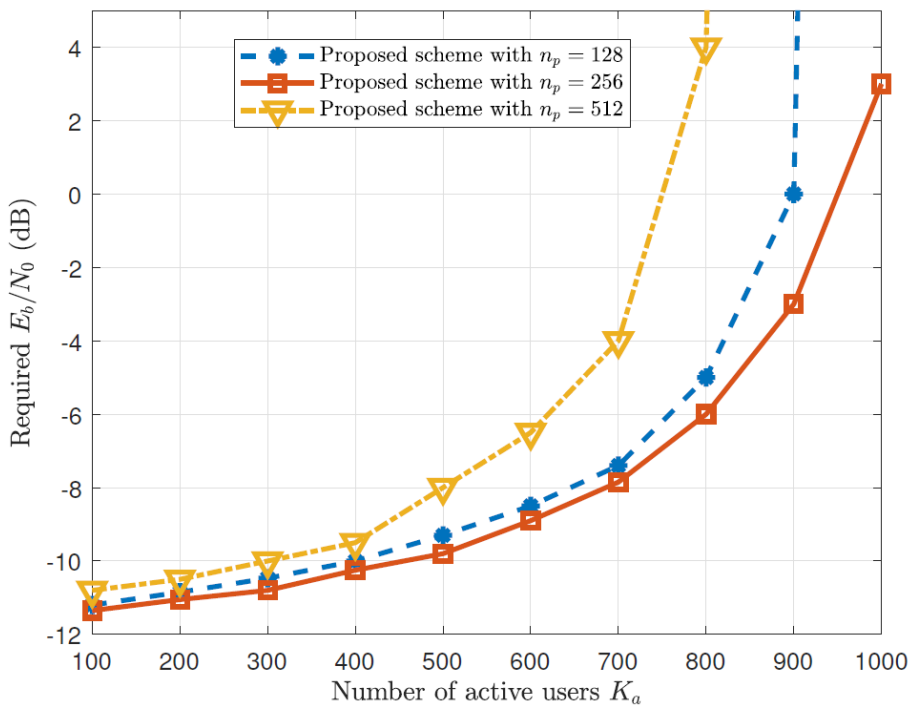


Figure 4.11: The required $E_b/N_0$ as a function of the number of active users in the proposed scheme for $M = 50$, $J = 2$, and different values of $n_p$.

To show the effect of selecting unequal lengths for different pilot parts, we have plotted the $E_b/N_0$ of MS-MRA for different values of $n_{p1}$ and $n_{p2}$ in Figure 4.12. It can be seen that the optimal choice is to select equal lengths for all pilot parts, i.e., $n_{p1} = n_{p2}$.

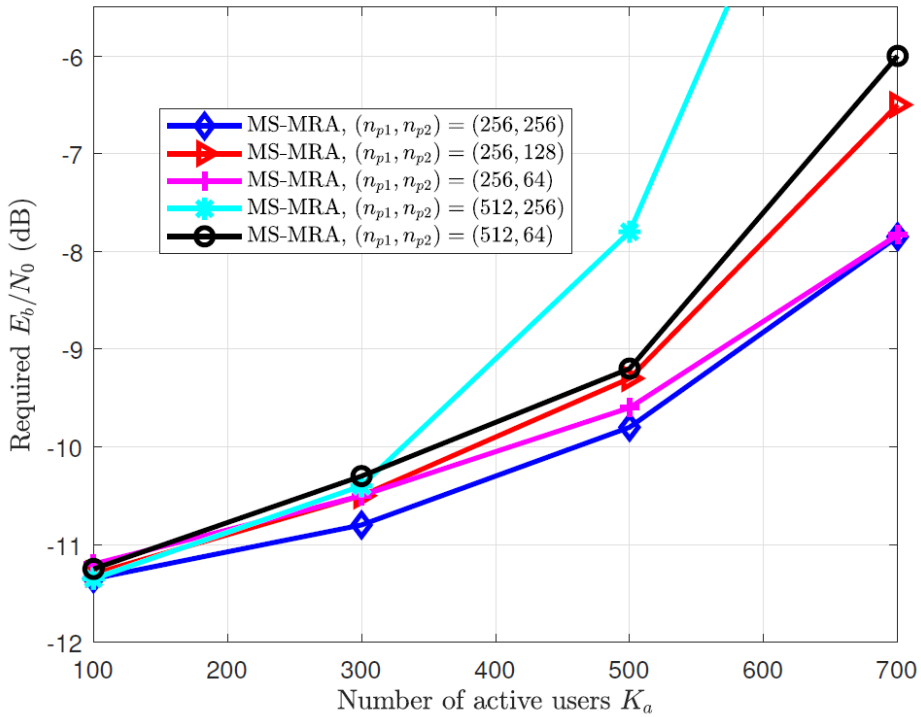


Figure 4.12: The required $E_b/N_0$ of MS-MRA for $J = 2$, and different sizes of pilots.

## 4.5 Chapter Summary

We propose a family of unsourced random access solutions for MIMO Rayleigh block fading channels. The proposed approaches employ a slotted structure with multiple stages of orthogonal pilots. The use of a slotted structure along with the orthogonal pilots leads to the lower computational complexity at the receiver, and also makes the proposed designs implementable for fast fading scenarios. We further improve the performance of the proposed solutions when the number of active users is very large by randomly dividing the users into different interleaver-power groups. The results show that the proposed MIMO URA designs are superior for both short and large blocklengths, while offering a lower computational complexity.

# Chapter 5

# RISUMA: RIS-Aided Unsourced Multiple Access

In this chapter, we consider a URA set-up equipped with a passive RIS, where a massive number of unidentified users (only a small fraction of them being active at any given time) transmit their data to the base station BS, without any collaborations among themselves or with the BS. We introduce a slotted coding scheme where each user chooses a slot at random for transmitting its signal, which consists of a pilot part and a randomly spread polar codeword. The proposed decoder operates in two phases. In the first phase, called the RIS configuration phase, the BS detects the pilots transmitted by users. The detected pilots are then utilized to estimate their corresponding channel state information, using which the BS suitably selects RIS phase shift employing the proposed RIS design algorithms, namely, alternating SDR (ASDR) and adaptive eigenvalue decomposition (AEVD). The proposed channel estimator offers the capability to estimate the channel coefficients of the users whose pilots interfere with each other without prior access to the list of transmitted pilots or the number of active users. In the second phase, called the data phase, transmitted messages of active users are decoded. It is illustrated that in the scenarios where the direct user-BS links are completely blocked or significantly attenuated, employing RIS improves the performance of a URA system by creating additional links between the BS and the

users. The proposed scheme outperforms the current state-of-the-art RIS-aided URA schemes, demonstrating up to a 6 dB improvement in the required energy.

The chapter is organized as follows. An introduction is provided in Section 5.1. In Section 5.2, we describe the system model. Sections 5.3 and 5.4 present the proposed RIS-aided URA scheme, taking into account the blocked user-BS channel in Section 5.3 and the existing user-BS channel in Section 5.4. We also describe the proposed encoder, pilot detector, channel estimator, RIS design, and decoding procedures in detail. We provide simulation results in Section 5.5, and conclude the chapter in Section 5.6.

## 5.1 Introduction

Reconfigurable intelligent surface is a promising technology developed for providing high spectral efficiency and energy savings for 5G and beyond wireless communication systems. Specifically, a passive RIS equipped with many low-cost passive elements, which can intelligently tune the phase shift of the incident electromagnetic waves, and reflect them in a desired direction without any amplification, can improve the efficiency of the network by enabling line-of-sight paths between the transmitters and the receivers in problematic environments with many blocking obstacles [34, 44, 45, 97].

URA schemes in the literature consider direct links between all the users and the BS [35–38,64]; however, in certain environments, the direct link between some users and the BS may be blocked or significantly attenuated. Therefore, the use of RIS can improve user connectivity in the URA by creating high-quality links between the BS and the users.

One of the most crucial problems in RIS-aided communication systems is the design of RIS reflecting coefficients to achieve the best system performance. This is solved via various approaches such as alternating optimization [44, 92], Gaussian randomization [34, 93], gradient descent [94], and semidefinite relaxation

(SDR) [45,97]. To design the RIS elements, a reliable knowledge of channel state information (CSI) is required. For this purpose, many grant-based multiple-access schemes solve the channel estimation problem by providing each user with an individual time slot [46,48–51,95]. Specifically, they assign each user an orthogonal pilot so that its CSI can be estimated without suffering from interference due to the simultaneous transmission of other users. However, orthogonal transmission is not feasible in the URA, hence new solutions are needed. Authors in [34] employ passive RIS to improve the URA system, which performs joint channel estimation and data detection in the presence of interfering signals of several users, all without the need for pilot transmissions. Every user transmits a rank-1 tensor, and as a result, the received signal is a tensor with a rank equal to the number of active users, perturbed by AWGN noise. Utilizing a coupled tensor decomposition technique at the receiver, the signals from distinct users along with their respective channel coefficient vectors are jointly estimated. In addition to channel estimation and data detection, they also propose a RIS design algorithm. Despite a suitable performance in the case of Rayleigh RIS-BS channel with full-rank channel coefficient matrix, this study encounters notable performance deterioration in cases where rank-deficient RIS-BS channel matrices are considered.

In this chapter, we propose a slotted URA scheme facilitated by a passive RIS coupled with the necessary channel estimation, RIS design, and pilot and data detection algorithms. In the proposed RIS-aided unsourced multiple access scheme (RISUMA), every user transmits a signal consisting of a pilot, appended to a polar codeword. The decoding process at the receiver takes place in two phases: the RIS configuration phase and the data phase. The RIS configuration phase is responsible for jointly detecting the active pilots and estimating the CSI via the newly proposed joint pilot detection and channel estimation (JDCE) algorithm, as well as designing the reflection coefficients of the RIS elements. During the data phase, the actual data transmission takes place. As stated above, unlike the grant-based schemes, where each identified user's CSI is estimated without any interference from the other users, in the URA, pilots from multiple users interfere with each other due to the unsourced nature of transmission for which there is

no cooperation between the users and the BS, or among the users themselves. Therefore, the CSI estimation task becomes more challenging. Nevertheless, the proposed JDCE algorithm is able to perform channel estimation successfully without any prior knowledge of the number of active users or their transmitted pilots. In the proposed RIS-aided URA set-up, in addition, we need to design the RIS reflection coefficients, which is not an issue in the standard URA. In particular, we propose two RIS design algorithms: alternating SDR (ASDR) and adaptive eigenvalue decomposition (AEVD). In the former, we alternatively solve a standard SDR problem, and update the RIS coefficients at each iteration. At each iteration of the latter algorithm, we employ the eigenvalue decomposition to find the appropriate RIS coefficients of each active user, and then the resulting coefficients of different users are combined. We show that the AEVD algorithm performs similarly to the ASDR, while having a lower computational complexity. On the other hand, for the encoder and decoder designs, similar solutions with other URA schemes can be adopted [37, 38]. Moreover, in this chapter, we employ polar codes along with SIC for the data phase to recover the transmitted messages. The aforementioned algorithms are devised for the situation where the direct communication links between the user and the BS are completely blocked. We also extend these algorithms to the scenario where there exist direct user-BS paths. We demonstrate that the newly proposed RISUMA method surpasses the CTAD algorithm introduced in [34], which currently stands as the state-of-the-art within RIS-assisted URA schemes.

Our contributions are as follows:

- We propose a RIS-assisted URA scheme, namely RISUMA, including appropriate encoding/decoding blocks, considering the direct user-BS link to be completely blocked. The proposed scheme offers superior performance compared to the state-of-the-art.
- A joint pilot detector and channel estimator algorithm (called JDCE) is proposed, which detects active pilots and estimates their corresponding channel coefficients in the pool of interfering signals without having knowledge about the number of active users and their identity. The proposed JDCE's

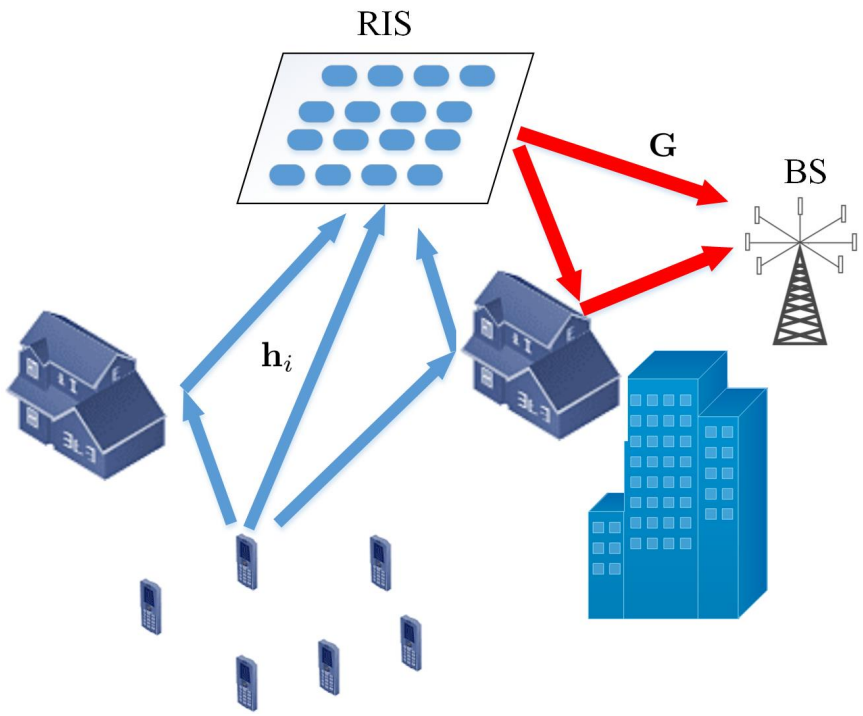


Figure 5.1: Illustration of a RIS-aided URA system.

ability for detection and channel estimation in the presence of interference makes it a unique and novel algorithm in the RIS literature.

- Two RIS phase shift design algorithms (namely, ASDR and AEVD) are devised whose goal is to increase the SINR of the input to the polar decoder, which is a suitable metric in URA systems.

- We modify the proposed decoding blocks to apply RISUMA to the scenario where a direct link exists between the users and the BS. In the resulting scheme, the direct user-BS channels are estimated as well as the cascaded channels. Then, the RIS design algorithm and the decoder are designed based on the overall channel coefficients of all the users.

## 5.2 System Model

Consider a RIS-aided URA system as depicted in Figure 5.1, where a BS with $M$ receive antennas serves $K_T$ single-antenna users (of which only $K_a$ are active in any transmission frame) with the help of a passive $N$-element RIS. We assume that both the BS and the RIS are equipped with a uniform planer array (UPA) [95]. Active users send $B$ bits of information to the BS through $n$ channel uses. Employing the Saleh-Valenzuela channel model, the RIS-BS channel is written as [34]

$$\mathbf{G} = \sqrt{MN} \sum_{l=1}^{L_G} \mu_l \mathbf{a}_M(\phi_{r,l}, \psi_{r,l})^T \mathbf{a}_N(\phi_{t,l}, \psi_{t,l}) \in \mathbb{C}^{M \times N}, \tag{5.1}$$

where $(\phi_{r,l}, \psi_{r,l})$ are the azimuth and elevation angles of arrival (AOA) at the BS from the $l$th path, $(\phi_{t,l}, \psi_{t,l})$ are the azimuth and elevation angles of departure (AOD) from the RIS to the BS through the $l$th path, $L_G$ is the number of paths from the RIS to the BS, $\mu_l$ denotes the $l$th path's gain which is modeled as a circularly symmetric complex Gaussian random variable, i.e., $\mu_l \sim \mathcal{CN}(0, L_0 d_l^{-\alpha_{\mathrm{PL}}})$ [96], $d_l$ is the length of the $l$th path, $L_0$ and $\alpha_{\mathrm{PL}}$ denote the path loss at the reference distance and the path loss exponent, respectively, and $\mathbf{a}_N(\phi, \psi)$ is the array steering vector of an $N_1 \times N_2$ UPA ($N = N_1 N_2$) which is represented as [1]

$$\mathbf{a}_N(\phi, \psi) = \bar{\mathbf{a}}_{N,\bar{\phi},\bar{\psi}} = \frac{1}{\sqrt{N}} e^{-j2\pi\bar{\phi}\mathbf{n}_1} \otimes e^{-j2\pi\bar{\psi}\mathbf{n}_2} \in \mathbb{C}^{1 \times N}, \tag{5.2}$$

where $\mathbf{n}_1 = \frac{d}{\lambda}[0, ..., N_1 - 1]$, $\mathbf{n}_2 = \frac{d}{\lambda}[0, ..., N_2 - 1]$, $\bar{\phi} = \sin(\phi)\cos(\psi)$, $\bar{\psi} = \sin(\psi)$, $\lambda$ denotes the carrier wavelength, and $d$ is the antenna spacing. Note also that $\mathbf{a}_M(\phi, \psi)$ is the array steering vector of an $M_1 \times M_2$ UPA ($M = M_1 M_2$) which is obtained by replacing $N$ by $M$ in (5.2). Similarly, the channel from the $i$th user to the RIS can be expressed as

$$\mathbf{h}_i = \sqrt{N} \sum_{f_i=1}^{L_{R,i}} \mu_{f_i} \mathbf{a}_N(\phi_{i,f_i}, \psi_{i,f_i}) \in \mathbb{C}^{1 \times N}, \tag{5.3}$$

[1] For the vector $\mathbf{t}$, we denote the element-wise exponentiation by $e^{\mathbf{t}}$.

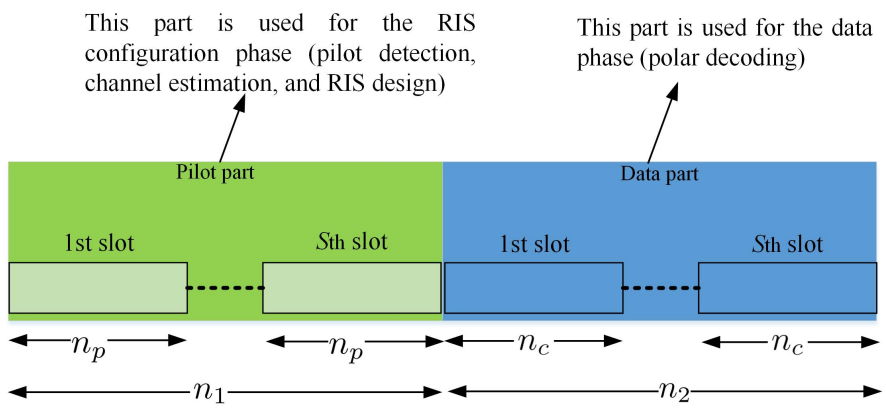


Figure 5.2: Transmission structure of the proposed URA scheme.

where $L_{R,i}$ is the number of paths between the $i$th user and the RIS, $(\phi_{i,f_i}, \psi_{i,f_i})$ are azimuth and elevation AOAs at the RIS for the $f_i$th path of the $i$th user's signal, with the distance $d_{f_i}$ and the corresponding path gain $\mu_{f_i}$(the path-loss model is the same as the one for the RIS-BS channel). The channel between the RIS and the BS is assumed to be perfectly known. This can be justified by the fact that the RIS and the BS are stationary, so the channel between them changes very slowly, and it can be well estimated with negligible overhead. Therefore, it is enough to focus on the estimation of the channel between each user and the RIS for our set-up.

## 5.3 RISUMA with Blocked User-BS link

Let $\mathbf{y}_t \in \mathbb{C}^{M\times 1}$ denote the uplink signal received at the BS at time $t$. Assuming that the direct link between the users and the BS is completely blocked, and the channel coefficients are constant throughout the frame, the received signal can be written as [95]

$$\mathbf{y}_t = \sum_{i=1}^{K_a} \mathbf{G}\text{diag}(\mathbf{h}_i)\mathbf{w}_t x_{i,t} + \mathbf{z}_t, \in \mathbb{C}^{M\times 1},\ t = 1, ..., n, \tag{5.4}$$

where $\mathbf{w}_t \in \mathbb{C}^{N\times 1}$ denotes the reflection coefficient vector of the RIS at time $t$ with $|[\mathbf{w}_t]_i| = 1$ (passive RIS assumption), $x_{i,t} \in \mathbb{C}$ is the symbol transmitted by

the $i$th user at time $t$, and $\mathbf{z}_t \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2 \mathbf{I}_M)$ represents the noise vector at time $t$.

The proposed coding scheme divides the entire frame into slot pairs. Each active user randomly selects only one slot pair to transmit its pilot and data signals as illustrated in Figure 5.2. At the receiver side, transmitted messages are decoded in two phases: the RIS configuration phase and the data phase. In the RIS configuration phase, active pilots in each slot are identified, and the corresponding CSI between the users and the RIS is estimated. Given the estimated channel of each active user, the BS designs the RIS reflection coefficients, and transmits them back to the RIS unit. After the phase shifts of the RIS elements are adjusted accordingly, the active users transmit their encoded signals in the data transmission phase. We describe the proposed encoder and decoder structures in more detail in the following.

### 5.3.1 Encoder

As shown in Figure 5.2, in the proposed transmission scheme, a length-$n$ frame is divided into pilot and data parts with lengths $n_1$ and $n_2$, respectively, where $n_2 + n_1 = n$. We also divide each part into $S$ slots of lengths $n_p = n_1/S$ and $n_c = n_2/S$. The $i$th user divides its $B$ bits of information into two parts: pilot part $\omega_{p_i} \in \{0,1\}^{1\times B_p}$ and data part $\omega_{c_i} \in \{0,1\}^{1\times B_c}$, where $B = B_p + B_c$. Every user obtains a preamble $\mathbf{b}_i \in \mathbb{C}^{1\times n_s}$ and a pilot $\mathbf{p}_i \in \mathbb{C}^{1\times n_p}$ by mapping $\omega_{p_i}$ to the rows of the preamble codebook $\mathbf{B} \in \mathbb{C}^{2^{B_p}\times n_s}$ and the pilot codebook $\mathbf{P} \in \mathbb{C}^{2^{B_p}\times n_p}$, respectively. The $i$th user adds cyclic redundancy check (CRC) bits to $\omega_{c_i}$, and encodes the result with an $(n_d,\ B_c + r)$ polar code, where $r$ is the number of CRC bits. The encoded bits are then modulated using binary phase shift keying (BPSK), which results in $\mathbf{v}_i = [v_{i,1}, v_{i,2}, ...v_{i,n_d}] \in \{\pm 1\}^{1\times n_d}$. To construct the signal of the data part, the encoder spreads each symbol of the modulated signal by its preamble [36,64], i.e., $\mathbf{c}_i = \mathbf{v}_i \otimes \mathbf{b}_i \in \mathbb{C}^{1\times n_c}$, where $n_c = n_s n_d$. The $i$th user randomly selects a slot index (denoted by $\zeta_i$), and transmits $\mathbf{p}_i$ and $\mathbf{c}_i$ through the $\zeta_i$th slot of the pilot and data parts, respectively.

Using the signal model in (5.4), the received signal at the $s$th slot of the pilot part can be written as

$$\mathbf{Y}_p = \sqrt{P_p} \sum_{i \in \mathcal{S}_s} \mathbf{G}\text{diag}(\mathbf{h}_i)\mathbf{W}_{p_s}\text{diag}(\mathbf{p}_i) + \mathbf{Z}_p \in \mathbb{C}^{M \times n_p}, \tag{5.5}$$

where $P_p$ denotes the signal power in the pilot part, $\mathcal{S}_s$ is the set of users in the $s$th slot, $\mathbf{W}_{p_s} = [\mathbf{w}_{j_{1,s}}, ..., \mathbf{w}_{j_{n_p,s}}]$ contains the RIS reflecting coefficients in the $s$th slot of the pilot part, $\mathbf{Z}_p = [\mathbf{z}_{j_{1,s}}, ..., \mathbf{z}_{j_{n_p,s}}]$, and $j_{t,s} = (s-1)n_p + t$. We omit the slot index from matrices for notational simplicity. Similarly, the received signal at the $s$th slot of the data part corresponding to the $f$th ($f = 1, 2, ..., n_c$) symbol of the modulated signal can be written as

$$\mathbf{Y}_{c,f} = \sqrt{P_c} \sum_{i \in \mathcal{S}_s} \mathbf{G}\text{diag}(\mathbf{h}_i)\mathbf{W}_{c_s}\text{diag}(\mathbf{b}_i)v_{i,f} + \mathbf{Z}_{c,f} \in \mathbb{C}^{M \times n_s}, \tag{5.6}$$

where $P_c$ denotes the signal power in the data part, $\mathbf{W}_{c_s} \in \mathbb{C}^{N \times n_s}$ contains the RIS reflecting coefficients in the $s$th slot of the data part, $\mathbf{Z}_{c,f} = [\mathbf{z}_{k_{1,s}}, ..., \mathbf{z}_{k_{n_s,s}}]$, and $k_{t,s} = Sn_p + (s-1)n_c + (f-1)n_s + t$. Note that we consider two choices for selecting $\mathbf{W}_{c_s}$. $\mathcal{C}_0$: the RIS coefficient vectors remain constant during each symbol's duration, and $\mathcal{C}_1$: the RIS coefficient vectors vary symbol by symbol, i.e.,

$$\mathbf{W}_{c_s} = \begin{cases} [\mathbf{w}_{f,s}, ..., \mathbf{w}_{f,s}] & \mathcal{C}_0 \\ [\mathbf{w}_{k_{1,s}}, ..., \mathbf{w}_{k_{n_s,s}}] & \mathcal{C}_1 \end{cases}. \tag{5.7}$$

### 5.3.2 Decoder

#### 5.3.2.1 Joint Pilot Detection and Channel Estimation (JDCE)

In the following, we describe the steps of the JDCE algorithm. It is clear from (5.3) that detecting all the active paths (finding angles of active paths), and estimating the corresponding path gains give us an estimate of the user-RIS

channel $\mathbf{h}_i$. The pilot with the highest probability of existence is detected first, and then the strongest path over which the detected pilot arrives at the RIS is identified. We also estimate the gain corresponding to the detected path using the least-squares (LS) technique. Finally, we remove the contribution of the detected pilot-path pair from the received signal, and repeat the procedure on the remaining signal to detect the next pilot-path pair. Note that since the residual received signal is updated at each iteration, we denote it by $\mathbf{Y}_p^j$, where $j$ is the iteration index. Details of the proposed JDCE algorithm are given below:

- **Step 1** (pilot detection): the pilot with the highest probability of existence is detected using the energy detection approach as in [36, 64]

$$\hat{k} = \underset{k \in 1,2,\ldots,2^{B_p}}{\arg\max} \frac{\text{trace}(\mathbf{Q}_k \mathbf{R}_{pj} \mathbf{Q}_k^H)}{\text{trace}(\mathbf{Q}_k \mathbf{Q}_k^H)}, \tag{5.8}$$

  where $\mathbf{R}_{pj} = \mathbf{Y}_p^{jH} \mathbf{Y}_p^j$, $\mathbf{Q}_k = \mathbf{W}_{p_s} \text{diag}(\bar{\mathbf{p}}_k)$, and $\bar{\mathbf{p}}_k$ is the $k$th row of the codebook $\mathbf{P}$.

- **Step 2** (path detection): By solving the following problem, we find the strongest path through which the $\hat{k}$th pilot (the detected pilot) arrives at the RIS

$$(\hat{l}, \hat{q}) = \underset{l \in \mathcal{T}(N_1), q \in \mathcal{T}(N_2)}{\arg\max} \frac{\text{trace}(\mathbf{F}_{l,q} \mathbf{R}_{pj} \mathbf{F}_{l,q}^H)}{\text{trace}(\mathbf{F}_{l,q} \mathbf{F}_{l,q}^H)}, \tag{5.9}$$

  where $\mathbf{F}_{l,q} = \mathbf{G} \text{diag}(\bar{\mathbf{a}}_{N,l,q}) \mathbf{W}_{p_s} \text{diag}(\bar{\mathbf{p}}_{\hat{k}})$, and $\bar{\mathbf{a}}_{N,l,q}$ is defined in (5.2).

- **Step 3** (SIC): from (5.5), we can infer that the contribution of the received signal corresponding to the detected pilot-path pair in the $j$th iteration can be obtained as

$$\mathbf{U}_j = \sqrt{P_p} \mathbf{G} \text{diag}(\bar{\mathbf{a}}_{N,\hat{l},\hat{q}}) \mathbf{W}_{p_s} \text{diag}(\bar{\mathbf{p}}_{\hat{k}}) \in \mathbb{C}^{M \times n_p}. \tag{5.10}$$

  Vectorizing both sides of (5.5), the received signal vector at the initial iteration ($\mathbf{y}_p^1 = \text{vec}(\mathbf{Y}_p^1) \in \mathbb{C}^{M n_p \times 1}$) can be written as

$$\mathbf{y}_p^1 = \sum_{j'=1}^{j} \bar{\mu}_{j'} \mathbf{u}_{j'} + \mathbf{z}_p', \tag{5.11}$$

where $\mathbf{u}_j = \text{vec}(\mathbf{U}_j) \in \mathbb{C}^{Mn_p \times 1}$, $\bar{\mu}_{j'} \in \mathbb{C}$ is the path gain corresponding to the pilot-path pair detected at the $j'$th iteration, and $\mathbf{z}'_p \in \mathbb{C}^{Mn_p \times 1}$ contains the signals corresponding to the pilot-path pairs that are not detected yet along with the vectorized noise term. The path gains are estimated using LS as

$$\hat{\mathbf{m}} = (\bar{\mathbf{U}}^H \bar{\mathbf{U}})^{-1} \bar{\mathbf{U}}^H \mathbf{y}_p^1, \tag{5.12}$$

where the $k$th column of $\bar{\mathbf{U}} \in \mathbb{C}^{Mn_p \times j}$ is $\mathbf{u}_k$. Then, the contribution of $\bar{\mathbf{U}}$ is removed from the initially received signal to obtain

$$\mathbf{y}_p^{(j+1)} = \mathbf{y}_p^1 - \bar{\mathbf{U}}\hat{\mathbf{m}}. \tag{5.13}$$

Finally, $\mathbf{y}_p^{(j+1)}$ is converted back to an $M \times n_p$ matrix $\mathbf{Y}_p^{(j+1)}$ before being passed to Step 1 for the $(j+1)$th iteration. The procedure is stopped if the condition

$$\frac{1}{Mn_p} \frac{\|\mathbf{y}_p^{(j+1)}\|^2 - \|\mathbf{y}_p^j\|^2}{\sigma_z^2} \leq \alpha_1 \tag{5.14}$$

is satisfied, where $\alpha_1 \in \mathbb{R}^+$ is a threshold.

#### 5.3.2.2 Data Phase

We now devise an iterative algorithm to detect the message bits $\omega_{c_i}$ using the received signal corresponding to the data part. The algorithm 1) generates log-likelihood ratios (LLRs) of the polar coded bits using the estimated channel coefficients, 2) passes the LLRs to the polar list decoder, 3) removes the contribution of the successfully decoded users (users whose detected messages satisfy the CRC check) from the received signal. The remaining received signal is passed back to the first step to generate a new LLR. Details of the proposed algorithm are given below.

We can rewrite the received signal in (5.6) as

$$\mathbf{Y}_{c,f} = \sum_{i \in \mathcal{S}_s} \mathbf{T}_i v_{i,f} + \mathbf{Z}_{c,f}, \tag{5.15}$$

where $\mathbf{T}_i = \sqrt{P_c}\mathbf{G}\text{diag}(\mathbf{h}_i)\mathbf{W}_{c_s}\text{diag}(\mathbf{b}_i) \in \mathbb{C}^{M\times n_s}$. Vectorizing both sides of (5.15), we write

$$\begin{aligned}\mathbf{y}_{c,f} &= \sum_{i\in\mathcal{S}_s}\mathbf{t}_i v_{i,f} + \mathbf{z}_{c,f}\\ &= \bar{\mathbf{T}}\tilde{\mathbf{v}}_f + \mathbf{z}_{c,f},\end{aligned} \tag{5.16}$$

where

$$\begin{aligned}\mathbf{t}_i &= \text{vec}(\mathbf{T}_i)\\ &= \sqrt{P_c}\text{vec}(\mathbf{G}\text{diag}(\mathbf{h}_i)\mathbf{W}_{c_s}\text{diag}(\mathbf{b}_i)) \in \mathbb{C}^{Mn_s\times 1},\end{aligned} \tag{5.17}$$

where $\mathbf{y}_{c,f} = \text{vec}(\mathbf{Y}_{c,f}) \in \mathbb{C}^{Mn_s\times 1}$, $\mathbf{z}_{c,f} = \text{vec}(\mathbf{Z}_{c,f}) \in \mathbb{C}^{Mn_s\times 1}$, $v_{i,f}$ is the $i$th row of $\tilde{\mathbf{v}}_f \in \mathbb{C}^{|\mathcal{S}_s|\times 1}$, and $\mathbf{t}_i$ is the $i$th column of $\bar{\mathbf{T}} \in \mathbb{C}^{Mn_s\times|\mathcal{S}_s|}$. We estimate $\tilde{\mathbf{v}}_f$ using the minimum mean square error (MMSE) estimator as

$$\hat{\mathbf{v}}_f = [\hat{v}_{1,f}, \hat{v}_{2,f}, ..., \hat{v}_{|\mathcal{S}_s|,f}]^T = \bar{\mathbf{T}}^H\hat{\mathbf{R}}^{-1}\mathbf{y}_{c,f}, \tag{5.18}$$

where

$$\hat{\mathbf{R}} = \mathbb{E}\{\mathbf{y}_{c,f}\mathbf{y}_{c,f}^H\} = \bar{\mathbf{T}}\bar{\mathbf{T}}^H + \sigma_z^2\mathbf{I}_{Mn_s}. \tag{5.19}$$

In a similar way as in [64], the LLR of the $f$th symbol of the $i$th user is approximated as

$$\hat{g}_{i,f} \approx 2\text{Re}\left(\hat{v}_{i,f}/\delta_i\right), \tag{5.20}$$

where $\delta_i$ is the $i$th diagonal element of the matrix $\mathbf{\Sigma} = \mathbf{I}_{|\mathcal{S}_s|} - \bar{\mathbf{T}}^H\hat{\mathbf{R}}^{-1}\bar{\mathbf{T}}$. The obtained LLR values are then passed to the polar list decoder. Finally, the contributions of users whose decoded message sequences satisfy the CRC check are removed from the received signal employing

$$\mathbf{Y}_{c,f} = \mathbf{Y}_{c,f} - \mathbf{T}_i\bar{v}_{i,f}, \tag{5.21}$$

where $\bar{v}_{i,f}$ is the $f$th symbol of the $i$th user obtained by encoding and modulating its successfully decoded message. The remaining signal $\mathbf{Y}_{c,f}$ is passed back to the MMSE estimator in (5.18) for the next iteration. The step terminates if no user is successfully decoded during an iteration.

#### 5.3.2.3 RIS Design

Unlike the RIS phase shift coefficient matrix $\mathbf{W}_{p_s}$, which is arbitrarily selected, in this section, we propose algorithms for designing the RIS phase shift matrix for use in the data part, denoted by $\mathbf{W}_{c_s}$. Note that since the RIS algorithm is designed based on the decoder, we introduce it here, after studying the data phase of the decoder. However, in practice, the RIS coefficients are obtained before data transmission.

We know from Section 5.3.2.2 that the MMSE estimate of the BPSK signal is fed to the polar decoder. Thus, improving the SINR at the output of the MMSE block is a good way to decrease the decoding error of each user. Plugging (5.16) into (5.18), we can obtain the MMSE estimate of the $f$th symbol of the $i$th user's codeword as

$$\hat{v}_{i,f} = \mathbf{t}_i^H \hat{\mathbf{R}}^{-1} \mathbf{t}_i v_{i,f} + \mathbf{t}_i^H \hat{\mathbf{R}}^{-1} \left( \sum_{k \in \mathcal{S}_s, k \neq i} \mathbf{t}_k v_{k,f} + \mathbf{z}_{c,f} \right), \tag{5.22}$$

where the first and second terms are signal and interference-plus-noise terms, respectively, whose powers can be computed as

$$\begin{aligned} \sigma_{s,i}^2 &= \mathbb{E}\{\|\mathbf{t}_i^H \hat{\mathbf{R}}^{-1} \mathbf{t}_i v_{i,f}\|^2\} \\ &= (\mathbf{t}_i^H \hat{\mathbf{R}}^{-1} \mathbf{t}_i)^2, \end{aligned} \tag{5.23}$$

$$\sigma_{\mathrm{IN},i}^2 = \mathbb{E}\{\|\hat{v}_{i,f}\|^2\} - \sigma_{s,i}^2 \tag{5.24}$$

$$\begin{aligned} &= \mathbb{E}\{\|\mathbf{t}_i^H \hat{\mathbf{R}}^{-1} \mathbf{y}_{c,f}\|^2\} - \sigma_{s,i}^2 \\ &= \mathbf{t}_i^H \hat{\mathbf{R}}^{-1} \mathbb{E}\{\mathbf{y}_{c,f} \mathbf{y}_{c,f}^H\} \hat{\mathbf{R}}^{-1} \mathbf{t}_i - \sigma_{s,i}^2 \\ &= \mathbf{t}_i^H \hat{\mathbf{R}}^{-1} \mathbf{t}_i - \sigma_{s,i}^2 \\ &= \sigma_{s,i} - \sigma_{s,i}^2. \end{aligned} \tag{5.25}$$

Therefore, the SINR of the $i$th user at the output of the MMSE estimator (input to the polar decoder) can be calculated as $\beta_i = \sigma_{s,i}/(1 - \sigma_{s,i})$. Using this SINR term, the decoding error probability of the $i$th user is obtained as [90]

$$e_{r,i} = \mathcal{F}(\sigma_{s,i}), \tag{5.26}$$

where

$$\mathcal{F}(x) = Q\left(\frac{0.5\log_2(1+x') - \dfrac{B_c + r}{n_d}}{\sqrt{\dfrac{1}{n_d}\dfrac{x'(x'+2)\log_2^2 e}{2(x'+1)^2}}}\right), \tag{5.27}$$

with $x' = x/(1-x)$, and $Q(.)$ denotes the standard $Q$-function. Using *vec trick* property [100], $\text{vec}(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3) = (\mathbf{A}_3^T \otimes \mathbf{A}_1)\text{vec}(\mathbf{A}_2)$, we can write $\mathbf{t}_i$ in (5.17) as

$$\mathbf{t}_i = \mathbf{E}_i\mathbf{w}_c, \tag{5.28}$$

where $\mathbf{E}_i$ and $\mathbf{w}_c$ are obtained as

$$\mathbf{E}_i = \begin{cases} \sqrt{P_c}\left(\mathbf{b}_i^T \otimes \mathbf{G}\text{diag}(\mathbf{h}_i)\right) \in \mathbb{C}^{Mn_s \times N} & \mathcal{C}_0 \\ \sqrt{P_c}\left(\text{diag}(\mathbf{b}_i) \otimes \mathbf{G}\text{diag}(\mathbf{h}_i)\right) \in \mathbb{C}^{Mn_s \times Nn_s} & \mathcal{C}_1 \end{cases}, \tag{5.29}$$

and

$$\mathbf{w}_c = \begin{cases} [\mathbf{W}_{c_s}]_{(:,i)} \in \mathbb{C}^{N\times 1} & \mathcal{C}_0 \\ \text{vec}(\mathbf{W}_{c_s}) \in \mathbb{C}^{Nn_s \times 1} & \mathcal{C}_1 \end{cases}, \tag{5.30}$$

respectively. Hence

$$\sigma_{s,i} = \mathbf{w}_c^H\mathbf{C}_i(\mathbf{w}_c)\mathbf{w}_c, \tag{5.31}$$

where

$$\mathbf{C}_i(\mathbf{w}_c) = \mathbf{E}_i^H\hat{\mathbf{R}}^{-1}\mathbf{E}_i. \tag{5.32}$$

Plugging (5.31) into (5.26), the RIS reflecting matrix that minimizes the total decoding error of the system is obtained by solving the following optimization problem

$$\arg\min_{\mathbf{w}_c} \sum_{i\in\mathcal{S}_s} \mathcal{F}(\mathbf{w}_c^H\mathbf{G}_i\mathbf{w}_c), \tag{5.33a}$$

$$\text{s.t.} \quad |[\mathbf{w}_c]_n| = 1, \tag{5.33b}$$

$$\mathbf{G}_i = \mathbf{C}_i(\mathbf{w}_c),\ i \in \mathcal{S}_s, \tag{5.33c}$$

where $\mathbf{G}_i$ is an auxiliary parameter matrix. It is evident that neither the objective function in (5.33a) nor the constraint in (5.33b) are convex. Therefore, it

cannot be solved using the standard convex optimization solvers. However, in the following, we proceed to adapt and refine it prior to solving by two distinct algorithms, namely ASDR and AEVD. Details of these algorithms are delineated below.

**ASDR:** It can be proved that once $x$ surpasses a specific threshold, namely $\bar{\alpha}$, there is no further significant reduction in the value of $\mathcal{F}(x)$, and $\mathcal{F}(x)$ is a non-increasing function of $x$. Given this motivation, the problem in (5.33) can be approximated as

$$\arg\max_{\mathbf{w}_c} \ \mathbf{w}_c^H \left( \sum_{i \in \mathcal{S}_s} \mathbf{G}_i \right) \mathbf{w}_c, \tag{5.34a}$$

$$\text{s.t.} \ |[\mathbf{w}_c]_n| = 1,$$

$$\mathbf{w}_c^H \mathbf{G}_i \mathbf{w}_c \leq \bar{\alpha}, \quad i \in \mathcal{S}_s, \tag{5.34b}$$

$$\mathbf{G}_i = \mathbf{C}_i(\mathbf{w}_c), \ i \in \mathcal{S}_s. \tag{5.34c}$$

We solve this problem using a two-step alternating optimization method: In the first step, we estimate $\mathbf{G}_i$ by $\mathbf{G}_i = \mathbf{C}_i(\mathbf{w}_{n-1})$, where $\mathbf{w}_{n-1}$ is the estimated RIS vector in the $(n-1)$th iteration, and in the second step, we solve the problem (5.34) given $\mathbf{G}_i$. The details of the second step of the algorithm are as follows. We define $\bar{\mathbf{W}} = \mathbf{w}_c \mathbf{w}_c^H$, where $\bar{\mathbf{W}} \succeq 0$ and $\text{rank}(\bar{\mathbf{W}}) = 1$, so we can write $\mathbf{w}_c^H \mathbf{G}_i \mathbf{w}_c = \text{trace}(\bar{\mathbf{W}} \mathbf{G}_i)$. Since the rank-one constraint is non-convex, we relax it using SDR to obtain the following convex semidefinite program (SDP) from (5.34)

$$\arg\max_{\bar{\mathbf{W}}} \ \ \text{trace}(\bar{\mathbf{W}} \bar{\mathbf{G}}), \tag{5.35}$$

$$\text{s.t.} \ [\bar{\mathbf{W}}]_{(n,n)} = 1, \ n = 1, ..., Nn_s$$

$$\text{trace}(\bar{\mathbf{W}} \mathbf{G}_i) \leq \bar{\alpha}, i \in \mathcal{S}_s,$$

where $\bar{\mathbf{G}} = \sum_{i \in \mathcal{S}_s} \mathbf{G}_i$. The problem in (5.35) can be solved using convex optimization solvers such as CVX. Note that the solution to the relaxed problem (5.35) is not necessarily rank-1. Thus, we perform additional steps similar to [97] to ensure that the rank-1 constraint is satisfied. Particularly, by applying the eigenvalue decomposition (EVD) on the solution, $\bar{\mathbf{W}} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^H$, the sub-optimal

solution to (5.35) is obtained as

$$\hat{\mathbf{w}}_c = \underset{l=1,\ldots,T_{SDR}}{\arg\min} \sum_{i\in\mathcal{S}_s} \mathcal{F}(\tilde{\mathbf{w}}_l^H \mathbf{G}_i \tilde{\mathbf{w}}_l), \tag{5.36}$$

where $[\tilde{\mathbf{w}}_l]_i = [\bar{\mathbf{u}}_l]_i / |[\bar{\mathbf{u}}_l]_i|$, $\bar{\mathbf{u}}_l = \mathbf{U}\mathbf{\Sigma}^{1/2}\mathbf{r}_l$, $\mathbf{r}_l \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{Nn_s^t})$ with $t \in \{0,1\}$ corresponding to $\mathcal{C}_0$ and $\mathcal{C}_1$, respectively, and $T_{SDR}$ denotes the number of realizations of $\mathbf{r}_l$. Note that for finding the solution in (5.36), we minimize the main cost function in (5.33) instead of the approximated one in (5.34). The details of the ASDR method for $\mathcal{C}_t$ scenario are shown in Algorithm 3, where $T_{\text{iter}}$ is the total number of iterations.

**Algorithm 3:** ASDR method for the RIS design for $\mathcal{C}_t$

**Initialization**: $\mathbf{w}_0 = [e^{j\theta_1}, ..., e^{j\theta_{Nn_s^t}}]^T$, where $\theta_j \sim U(0, 2\pi)$
**for** $n = 1, 2, ..., T_{\text{iter}}$ **do**

1. Calculate $\mathbf{G}_i = \mathbf{C}_i(\mathbf{w}_{n-1})$ according to (5.32).
2. Calculate $\bar{\mathbf{W}}$ by SDP
3. Perform EVD, $\bar{\mathbf{W}} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H$.
4. Calculate $\mathbf{w}_n$ according to (5.36).

**end**

**Algorithm 4:** AEVD method for the RIS design for $\mathcal{C}_t$

**Initialization**: $\mathbf{w}_0 = [e^{j\theta_1}, ..., e^{j\theta_{Nn_s^t}}]^T$, where $\theta_j \sim U(0, 2\pi)$
**for** $n = 1, 2, ..., T_{\text{iter}}$ **do**
1) Calculate $\mathbf{G}_i = \mathbf{C}_i(\mathbf{w}_{n-1})$ according to (5.32).
2) Calculate $\mathbf{q}_{i,\max}$.
3) $\mathbf{v}_n = \dfrac{1}{|\mathcal{S}_s|} \sum_{i\in\mathcal{S}_s} \mathbf{q}_{i,\max}$.
4) $[\mathbf{w}_n]_i = [\mathbf{v}_n]_i / |[\mathbf{v}_n]_i|, i = 1, ..., Nn_s^t$
5) $P_e(n) = \sum_{i\in\mathcal{D}_s} \mathcal{F}(\mathbf{w}_n^H \mathbf{G}_i \mathbf{w}_n)$.
**if** $P_e(n) < \alpha_2$ **then**
Stop the algorithm.
**end**
**end**
**Output:** $\hat{\mathbf{w}}_c = \mathbf{w}_{n'}$, where $n' = \underset{n}{\text{argmin}}\ P_e(n)$

**AEVD:** To decrease the computational complexity of the RIS design algorithm, we propose AEVD for solving (5.33). The AEVD method comprises the following steps: 1) First, we identify the RIS coefficient vectors that minimize the error for each user individually. To achieve this, we compute the eigenvector of the matrix $\mathbf{G}_i$ corresponding to its largest eigenvalue, denoted as $\mathbf{q}_{i,\max}$[2], 2) subsequently, we calculate the average of these individual vectors to obtain a combined vector, 3) to ensure compliance with the unit-modulus constraint of the RIS elements, we rescale each element of the resulting vector. We repeat this AEVD algorithm for a total of $T_{\text{iter}}$ iterations, stopping only when the total decoding error falls below the predefined threshold $\alpha_2$. A detailed description of the AEVD algorithm can be found in Algorithm 4.

## 5.4 RISUMA with Direct User-BS Link

In this section, we extend the proposed RISUMA approach by modifying it to accommodate scenarios where a direct path exists between the users and the BS. In this case, the system model is the same as in Sections 5.2, except for the existence of direct paths between the users and the BS. Then, in a similar way as in (5.4), the received signal at the BS can be modeled as

$$\mathbf{y}_t = \sum_{i=1}^{K_a} \left(\mathbf{G}\text{diag}(\mathbf{h}_i)\mathbf{w}_t + \mathbf{d}_i\right) x_{i,t} + \mathbf{z}_t, \in \mathbb{C}^{M\times 1} \quad , \ t = 1, ..., n, \tag{5.37}$$

where $\mathbf{d}_i$ is modeled as

$$\mathbf{d}_i = \sqrt{M} \sum_{g_i=1}^{L_{B,i}} \mu_{g_i} \mathbf{a}_M(\phi_{i,g_i}, \psi_{i,g_i})^T \in \mathbb{C}^{M\times 1}, \tag{5.38}$$

where $L_{B,i}$ is the number of direct paths between the $i$th user and the BS, respectively, $(\phi_{i,g_i}, \psi_{i,g_i})$ are azimuth and elevation AOAs at the BS for the $g_i$th

[2]For a Hermitian matrix $\mathbf{C}$ and a vector with unit norm, $\|\mathbf{w}\|^2 = 1$, the maximum value of $c = \mathbf{w}^H\mathbf{C}\mathbf{w}$ is obtained by choosing $\mathbf{w}$ as the eigenvector of $\mathbf{C}$ corresponding to its largest eigenvalue [98]

direct path of the $i$th user's signal. For the user-BS direct path, we consider the same path-loss model as the ones presented for the RIS-BS and the user-RIS channels in Section 5.2 with $d_{g_i}$ and $\mu_{g_i}$ being the distance and path gain of the $g_i$th direct path, respectively. Employing the same encoder as in Section 5.3.1, while considering the user-BS direct channel, the received signal models in (5.5) and (5.6) are rewritten as

$$\mathbf{Y}_p = \sqrt{P_p} \sum_{i \in \mathcal{S}_s} \mathbf{G}\text{diag}(\mathbf{h}_i)\mathbf{W}_{p_s}\text{diag}(\mathbf{p}_i) + \mathbf{d}_i\mathbf{p}_i + \mathbf{Z}_p, \tag{5.39}$$

$$\mathbf{Y}_{c,f} = \sqrt{P_c} \sum_{i \in \mathcal{S}_s} \left(\mathbf{G}\text{diag}(\mathbf{h}_i)\mathbf{W}_{c_s}\text{diag}(\mathbf{b}_i) + \mathbf{d}_i\mathbf{b}_i\right) v_{i,f} + \mathbf{Z}_{c,f}. \tag{5.40}$$

The decoder is also presented below.

### 5.4.1 Decoder

Similar to Section 5.3.2, the decoder in the presence of the direct user-BS link operates in two phases: The RIS configuration phase, where pilot detection, channel estimation, and RIS design are performed; and the data phase where transmitted bit sequences are decoded. Decoding blocks are described below.

#### 5.4.1.1 Joint Pilot Detection and Channel Estimation

In this part, we modify the JDCE algorithm in Section 5.3.2.1 to estimate $\mathbf{d}_i$ as well. Different steps of the modified version are described below.

- **Step 1** (Pilot detection): the pilot with the highest probability of existence

(with index $\hat{k}$) is obtained as

$$\hat{k} = \max(\hat{k}_1, \hat{k}_2), \tag{5.41}$$

$$\hat{k}_1 = \underset{k \in 1,2,\ldots,2^{B_p}}{\arg\max} \frac{\text{trace}(\mathbf{Q}_k \mathbf{R}_{pj} \mathbf{Q}_k^H)}{\text{trace}(\mathbf{Q}_k \mathbf{Q}_k^H)}, \tag{5.42}$$

$$\hat{k}_2 = \underset{k \in 1,2,\ldots,2^{B_p}}{\arg\max} \frac{\bar{\mathbf{p}}_k \mathbf{R}_{pj} \bar{\mathbf{p}}_k^H}{\|\bar{\mathbf{p}}_k\|^2}. \tag{5.43}$$

Note that if $\hat{k}_1 > \hat{k}_2$, the strongest scatterer belongs to the user-RIS-BS channel, and it belongs to the user-BS direct channel otherwise.

- **Step 2** (path detection): using the following problem, we find the AOAs of the $\hat{k}$th pilot's strongest scatterer.

$$(\hat{l}, \hat{q}) = \underset{l \in \mathcal{T}(N_1), q \in \mathcal{T}(N_2)}{\arg\max} \delta_{l,q}, \tag{5.44}$$

$$\delta_{l,q} = \begin{cases} \dfrac{\text{trace}(\mathbf{F}_{l,q} \mathbf{R}_{pj} \mathbf{F}_{l,q}^H)}{\text{trace}(\mathbf{F}_{l,q} \mathbf{F}_{l,q}^H)} & \text{if} \quad \hat{k} = \hat{k}_1 \\ \bar{\mathbf{a}}_{N,l,q}^* \mathbf{Y}_p^j \bar{\mathbf{p}}_{\hat{k}}^H & \text{if} \quad \hat{k} = \hat{k}_2 \end{cases}. \tag{5.45}$$

- **Step 3** (SIC): replacing

$$[\bar{\mathbf{U}}]_{(:,j)} = \begin{cases} \sqrt{P_p} \mathbf{F}_{\hat{l},\hat{q}} & \text{if} \quad \hat{k} = \hat{k}_1 \\ \sqrt{P_p} \bar{\mathbf{a}}_{N,\hat{l},\hat{q}}^T \bar{\mathbf{p}}_{\hat{k}} & \text{if} \quad \hat{k} = \hat{k}_1 \end{cases} \tag{5.46}$$

with the one in Section 5.3.2.1, the SIC operation in the $j$th iteration can be performed as in (5.13). The stopping criterion is also the same as in (5.14).

#### 5.4.1.2 Data Detection

We employ the same MMSE-based decoder as in Section 5.3.2.2, except for modifying the matrix $\mathbf{T}_i$ as

$$\mathbf{T}_i = \sqrt{P_c} \left( \mathbf{G} \text{diag}(\mathbf{h}_i) \mathbf{W}_{c_s} \text{diag}(\mathbf{b}_i) + \mathbf{d}_i \mathbf{b}_i \right) \in \mathbb{C}^{M \times n_s}, \tag{5.47}$$

to incorporate the direct path.

#### 5.4.1.3 RIS Design

In a similar way as in Section 5.3.2.3, the SINR of the $i$th user in the input of the polar decoder is calculated as $\beta_i = \sigma_{s,i}/(1-\sigma_{s,i})$ where $\sigma_{s,i} = \mathbf{t}_i^H \hat{\mathbf{R}}^{-1}\mathbf{t}_i$, and $\mathbf{t}_i = \text{vec}\,(\mathbf{T}_i)$. Then, using (5.47), we can write $\mathbf{t}_i = \mathbf{E}_i\mathbf{w}_c + \mathbf{e}_i$, where $\mathbf{E}_i$ and $\mathbf{w}_c$ are defined in (5.29) and (5.30), respectively, and $\mathbf{e}_i = \text{vec}(\mathbf{d}_i\mathbf{b}_i)$. Therefore, we can write $\sigma_{s,i}$ in the following form

$$\begin{aligned}\sigma_{s,i} &= \mathbf{w}_c^H\mathbf{E}_i^H\hat{\mathbf{R}}^{-1}\mathbf{E}_i\mathbf{w}_c + \mathbf{w}_c^H\mathbf{E}_i^H\hat{\mathbf{R}}^{-1}\mathbf{e}_i \\ &\quad + \mathbf{e}_i^H\hat{\mathbf{R}}^{-1}\mathbf{E}_i\mathbf{w}_c + \mathbf{e}_i^H\hat{\mathbf{R}}^{-1}\mathbf{e}_i, \\ &= \bar{\mathbf{w}}_c^H\mathbf{C}_i'(\bar{\mathbf{w}}_c)\bar{\mathbf{w}}_c \end{aligned} \tag{5.48}$$

where

$$\mathbf{C}_i'(\bar{\mathbf{w}}_c) = \begin{bmatrix} \mathbf{E}_i^H\hat{\mathbf{R}}^{-1}\mathbf{E}_i & \mathbf{E}_i^H\hat{\mathbf{R}}^{-1}\mathbf{e}_i \\ \mathbf{e}_i^H\hat{\mathbf{R}}^{-1}\mathbf{E}_i & \mathbf{e}_i^H\hat{\mathbf{R}}^{-1}\mathbf{e}_i \end{bmatrix}, \tag{5.49}$$

$$\bar{\mathbf{w}}_c = \begin{bmatrix} \mathbf{w}_c \\ 1 \end{bmatrix}. \tag{5.50}$$

We can see that the parameter $\sigma_{s,i}$ has exactly the same structures in (5.48) and (5.31). With the same argument as in Section 5.3.2.3, the RIS reflecting matrix that minimizes the total decoding error of the system is obtained by solving the following optimization problem

$$\arg\min_{\bar{\mathbf{w}}_c} \sum_{i\in\mathcal{S}_s} \mathcal{F}(\bar{\mathbf{w}}_c^H\mathbf{G}_i\bar{\mathbf{w}}_c), \tag{5.51a}$$

$$\text{s.t.} \quad |[\bar{\mathbf{w}}_c]_n| = 1,\ n = 1, 2, ..., Nn_s^t \tag{5.51b}$$

$$[\bar{\mathbf{w}}_c]_{Nn_s^t+1} = 1, \tag{5.51c}$$

$$\mathbf{G}_i = \mathbf{C}_i'(\bar{\mathbf{w}}_c),\ i \in \mathcal{S}_s, \tag{5.51d}$$

where $\mathbf{G}_i$ is an auxiliary parameter matrix. The problem in (5.51) can be solved via modified ASDR and AEVD algorithms as described below.

**ASDR:** The modified ASDR algorithm is obtained by revising Algorithm 3. Particularly, in the first step of Algorithm 3, we calculate $\mathbf{G}_i = \mathbf{C}_i'(\mathbf{w}_{n-1})$

according to (5.49). For the SDP in the second step, we replace $Nn_s^t + 1$ with $Nn_s^t$ for $\mathcal{C}_t$ structure. For the fourth step, $\hat{\mathbf{w}}_c$ is obtained by

$$\hat{\mathbf{w}}_c = \underset{l=1,\ldots,T_{SDR}}{\arg\min} \sum_{i \in \mathcal{S}_s} \mathcal{F}\left( \begin{bmatrix} \tilde{\mathbf{w}}_l \\ 1 \end{bmatrix}^H \mathbf{G}_i \begin{bmatrix} \tilde{\mathbf{w}}_l \\ 1 \end{bmatrix} \right). \tag{5.52}$$

**AEVD:** The modified AEVD is the same as Algorithm 4, except replacing $Nn_s^t+1$ with $Nn_s^t$, and correcting the first and second steps as follows. In the first step, the $\mathbf{G}_i$ matrix is obtained by $\mathbf{G}_i = \mathbf{C}_i'(\bar{\mathbf{w}}_{n-1})$. In the second step, we find the eigenvector corresponding to the largest eigenvector, $\mathbf{q}_{i,\max} \in \mathbb{C}^{(Nn_s^t+1)\times 1}$, then we calculate $\mathbf{q}_{i,\max} = \frac{\mathbf{q}_{i,\max}}{[\mathbf{q}_{i,\max}]_{Nn_s^t+1}}$ to satisfy the constraint in (5.51c).

### 5.4.2 Computational Complexity

In this part, we calculate the computational complexity of the proposed RISUMA. Note that we consider the number of multiplications as a measure of computational complexity.

#### 5.4.2.1 Joint Pilot Detection and Channel Estimation

Energy detectors in (5.8) and (5.41) has the computational complexity of $\mathcal{O}\left(MNn_p 2^{B_p}\right)$, where $\mathcal{O}(.)$ is the standard big-O notation, denoting the order of complexity. The computational complexity of the path detectors in (5.9) and (5.44) is $\mathcal{O}\left(NM^2 n_p + MN^2 n_p\right)$. Note that for calculating the order of computational complexity of these blocks, we use the fact that trace($\mathbf{ABB}^H\mathbf{A}^H$) can be implemented with the computational complexity of order $\mathcal{O}(ABC)$, where $\mathbf{A}$ and $\mathbf{B}$ are $A \times B$ and $B \times C$ matrices. Performing the SIC in (5.13) at the $j$th iteration has the complexity of order $\mathcal{O}(j^3 + j^2 M n_p)$. Besides, the maximum iterations for performing JDCE is upper bounded by $T_a = K_a L_a / S$, where $L_a$ is the average number of paths between each user and RIS, and we consider $|\mathcal{S}_s| \approx K_a/S$. Therefore, the total computational complexity of JDCE in $S$ slots is then upper

bounded by

$$\mathcal{O}\left(ST_aMNn_p\left(2^{B_p}+M+N\right)+S\sum_{j=1}^{T_a}(j^3+j^2Mn_p)\right). \tag{5.53}$$

#### 5.4.2.2 Data Detection

The computational complexity for LLR generation and SIC in Sections 5.3.2.2 and 5.4.1.2 is in the order of $\mathcal{O}((Mn_s)^3+|\mathcal{S}_s|(Mn_s)^2)$, and the polar list decoder has a computational complexity of $\mathcal{O}(|\mathcal{S}_s|n_d\log n_d)$. Considering $|\mathcal{S}_s|\approx K_a/S$, the total computational complexity for the data detection part can be approximated as

$$\mathcal{O}\left(S(Mn_s)^3+K_aM^2n_s^2+K_an_d\log n_d\right). \tag{5.54}$$

#### 5.4.2.3 RIS Design

In this part, we calculate the computational complexity of ASDR and AEVD for the $\mathcal{C}_t$ scenario. Calculating $\hat{\mathbf{R}}^{-1}$ in both algorithms has the complexity of $\mathcal{O}((Mn_s)^3+|\mathcal{S}_s|(Mn_s)^2)$. Calculating $\mathbf{C}_i(\bar{\mathbf{w}}_c)$ and $\mathbf{C}_i'(\bar{\mathbf{w}}_c)$ in (5.32) and (5.49) have the computational complexity of $\mathcal{O}(|\mathcal{S}_s|MNn_s^{(2+t)}(M+N))$. Finding the $\mathbf{q}_{i,\max}$ in AEVD has the computational complexity of $\delta_{\text{AEVD}}=\mathcal{O}(|\mathcal{S}_s|N^2n_s^{2t})$ [98]. For performing SDR in ASDR algorithm, the computational complexity is $\delta_{\text{ASDR}}=\mathcal{O}\left(\max(|\mathcal{S}_s|,Nn_s^t)^4\sqrt{Nn_s^t}\log(1/\epsilon')\right)$, where $\epsilon'>0$ denotes the solution accuracy [101]. Considering $|\mathcal{S}_s|\approx K_a/S$, the total complexity of ASDR and AEVD algorithms (which is performed in $T_{\text{iter}}$ iterations) in $S$ slots can be approximated as

$$\mathcal{O}\left(ST_{\text{iter}}\left(M^3n_s^3+|\mathcal{S}_s|MNn_s^{(2+t)}(M+N)+\delta_{\text{alg}}\right)\right), \tag{5.55}$$

where alg $\in\{\text{ASDR},\text{AEVD}\}$. Looking at (5.55), it is clear that setting $t=0$ (selecting the structure $\mathcal{C}_0$ in (5.7)) significantly decreases the computational complexity of the RIS design algorithms. Moreover, the computational complexity of AEVD is lower than that of ASDR.

## 5.5 Numerical Results

In this section, we provide several numerical examples to demonstrate the performance of the proposed RIS-aided URA solution. In all the simulations, unless otherwise stated, we consider the user-BS link to be completely blocked, and we choose $\sigma_z^2 = -95\text{dBm}$, $B_c = 90$, $B_p = 10$, $S = 4$, $n_d = 256$, $n_s = 10$, $n_p = 440$, $r = 16$, $L_G = 2$, $L_{R,i} = 2, i = 1, ..., K_a$, $M_1 = M_2 = N_1 = N_2 = 8$, $\alpha_1 = 0.01$, $L_0 = 10^{-3}$, $\alpha_{\text{PL}} = 2.3$ (for user-RIS and RIS-BS paths), and $d = \lambda/2$. The distances of each user-RIS and each user-BS paths are drawn from $d_i \in U(200m, 300m)$ and $d_i \in U(250m, 350m)$, respectively, and it is set as $d_i = 100m$ for every RIS-BS path. We draw every element of $\mathbf{B}$, $\mathbf{P}$, and $\mathbf{W}_{p_s}$ from $\mathcal{CN}(0, 1)$, then each element of $\mathbf{W}_{p_s}$ is rescaled to have unit modulus, every row of $\mathbf{B}$ is normalized to have a norm of $n_s$, and rows of $\mathbf{P}$ are scaled to satisfy $\|\mathbf{W}_{p_s}\text{diag}(\bar{\mathbf{p}}_i)\|^2 = \|\mathbf{W}_{p_s}\text{diag}(\bar{\mathbf{p}}_j)\|^2$, $\frac{1}{2^{B_p}}\sum_{i=1}^{2^{B_p}}\|\bar{\mathbf{p}}_i\|^2 = n_p$ (for a completely blocked user-BS path), and $\|\bar{\mathbf{p}}_i\|^2 = \|\bar{\mathbf{p}}_j\|^2 = n_p$ (for non-blocked user-BS path) with $\bar{\mathbf{p}}_i$ being the $i$th row of $\mathbf{P}$. To construct the steering vectors of each path for the RIS in (5.1) and (5.3), we randomly choose $\bar{\phi}$ and $\bar{\psi}$ from $\mathcal{T}(N_1)$ and $\mathcal{T}(N_2)$, respectively, where $\mathcal{T}(.)$ is as defined in the last paragraph of the introduction section. Similarly, the AOAs of each received path at the BS in (5.1) and (5.38) are randomly selected from the sets $\mathcal{T}(M_1)$ and $\mathcal{T}(M_2)$. The energy-per-bit and the PUPE of the system are defined as

$$E_b/N_0 = \frac{Pn_T}{B} \tag{5.56}$$

$$P_e = p_{fa} + p_{md}, \tag{5.57}$$

where $P$ and $n_T$ are the average per-symbol power and the total length of the transmitted signal of each user, $p_{md}$ is the probability that an active user's message is not decoded, and $p_{fa}$ is the probability that a decoded message is indeed not sent.

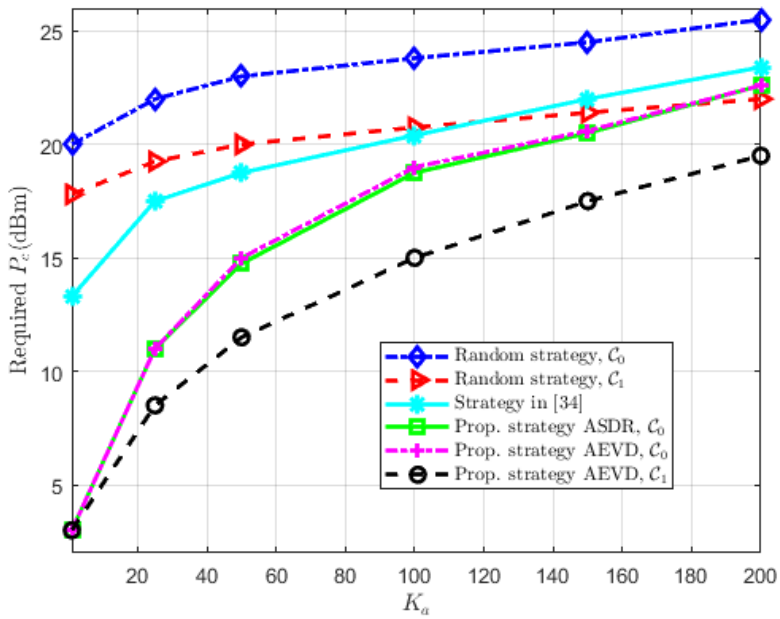


Figure 5.3: The required $P_c$ to achieve the target PUPE of 0.1 for $n_s = 10$, $n_d = 256$, and different RIS phase shift strategies.

In Figure 5.3, we depict the average transmit power required by the proposed scheme to achieve the target PUPE of 0.1 for different RIS design strategies and different RIS reflection strategies defined in (5.7). Note that the simulations are run for known CSI. We can deduce from this figure that the proposed beamforming strategy performs better than the strategy employed in [34] and the case of randomly generated phase shifts (resulting in up to 11dBm and 17dBm power savings, respectively). This improvement is due to the employment of a more suitable metric for the RIS design in the proposed algorithm, where the overall phase shift matrix is obtained by minimizing the decoding error probability. This is in contrast to the RIS design algorithm in [34] maximizes the minimum channel gain among active users. Besides, it is clear that the proposed AEVD performs comparable to the proposed ASVD (with $\bar{\alpha} = 0.53$), while having lower computational complexity. Also, we can interpret from this figure that employing $\mathcal{C}_1$ structure provides higher accuracy, however, according to our discussion in Section. 5.4.2, it suffers from a considerably higher computational complexity than the $\mathcal{C}_0$ case. Note that since we consider the direct link between each user and

the BS to be completely blocked in this simulation, no communication is possible without the help of RIS.

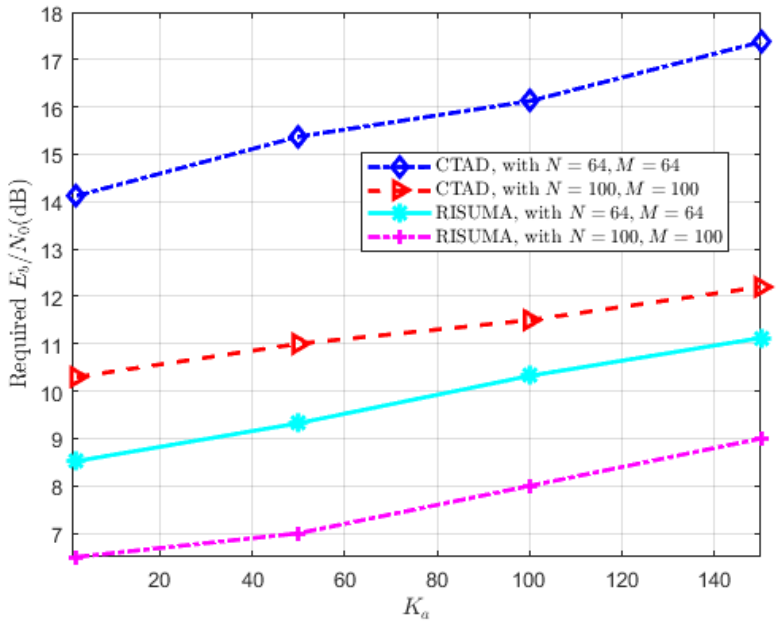


Figure 5.4: The required $E_b/N_0$ for the proposed RISUMA and CTAD in [34] for achieving a target PUPE of 0.1.

As shown in (5.1), RISUMA is designed based on the Saleh-Valenzuela model for the RIS-BS channel. However, to ensure a fair comparison with CTAD algorithm in [34], we consider a Rayleigh model assumption, i.e., the elements of RIS-BS channel matrix are generated as $[\mathbf{G}]_{(i,j)} \sim \mathcal{CN}(0, L_0 d_l^{\alpha_{\mathrm{PL}}})$, where $\mathrm{rank}(\mathbf{G}) = \min(M, N)$. The performance of the RISUMA and CTAD is compared in Figure 5.4 for $n = 12288$ ($S = 12$, $n_d = 512$, $n_s = 1$, $P_c/P_p = 0.5$, $n_p = 512$, and AEVD for RISUMA), $B = 316$, and different values of $M$ and $N$. We can see from this figure that the proposed RISUMA shows superior performance compared to CTAD. The reasons for this are: 1) RISUMA adopts a more reliable RIS design algorithm than CTAD, i.e., minimizing the decoding error of the polar code instead of maximizing the minimum channel gain of the users in CTAD, 2) RISUMA employs the powerful polar list decoder for decoding transmitted messages, while CTAD uses a dequantizer for mapping soft estimated

codewords to the bit sequences, 3) RISUMA employs SIC, which is an effective block in a URA set-up [37, 38]. We should also note that in the case of a rank-deficient $\mathbf{G}$ matrix ($\text{rank}(\mathbf{G}) < \min(M, N)$), there is a meaningful performance degradation in CTAD algorithm; however, the proposed RISUMA does not make any assumptions regarding the rank of this matrix, and is almost unaffected by it.

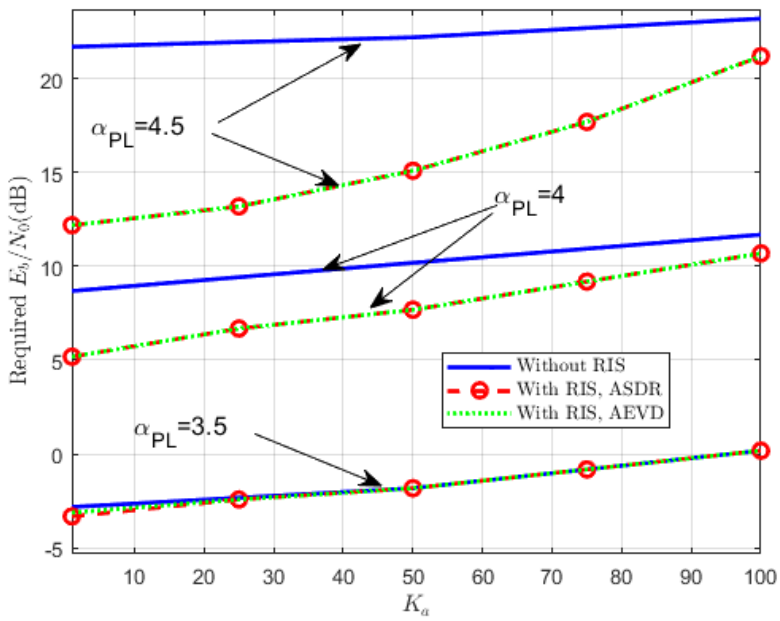


Figure 5.5: The required $E_b/N_0$ to achieve the target PUPE of 0.1 for different user-BS path-loss exponents.

To assess the contribution of the RIS to the efficiency of the URA system, in Figure 5.5, we compare the performance of the proposed RISUMA with and without employing RIS for achieving the target PUPE of 0.1, when the user-BS communication channel exists for different path-loss exponents, $\alpha_{\text{PL}} = 3.5, 4, 4.5, \infty$. Note that $\alpha_{\text{PL}} = \infty$ corresponds to the scenario of fully obstructed user-BS channels. We set $L_{B,i} = L_{R,i} = 1,\ i = 1, ..., K_a$, $\bar{\alpha} = 2$, $P_c/P_p = 0.01$, and employ the $\mathcal{C}_0$ strategy. It is shown in Figure 5.5 that for the stronger user-BS channel with path-loss exponent $\alpha_{\text{PL}} = 3.5$, employing RIS only slightly improves the performance of the system. However, for weaker user-BS channels (with the path-loss

exponents of the user-BS channel $\alpha_{\mathrm{PL}} = 4, 4.5$ and $\infty$), employing RIS improves the performance of the URA system considerably. Hence, employing RIS is crucial for scenarios with weak user-BS channels. It is seen that the AEVD operates as well as the ASDR algorithm, while having less computational complexity.

## 5.6 Chapter Summary

We have proposed a RIS-aided URA scheme to enable connectivity between the BS and the users whose direct channel to the BS is blocked or significantly attenuated. The proposed scheme operates in two phases: the RIS configuration phase and the data phase. In the former, transmitted pilots are identified in the presence of heavy interference, their corresponding CSI is estimated, and RIS reflection coefficients are suitably designed by employing two different approaches. In the data phase, the receiver detects the transmitted messages using a polar list decoder, and the contribution of successfully decoded messages is removed from the received signal using SIC. We have demonstrated that the proposed scheme enhances the URA system performance when the connection between users and the base station has significantly degraded. Moreover, it achieves a reduction in energy consumption of up to 4 dB when compared to the existing RIS-assisted URA algorithms.

# Chapter 6

# Summary and Conclusions

In this thesis, three different solutions are proposed for unsourced random access. The first algorithm utilizes random spreading of polar codewords, and proposes power diversity and SIC over GMACs as a way of increasing the number of supported users. The second URA scheme, designed for MIMO fading channels, leverages frame slotting, transmission of multiple stages of orthogonal pilots, the implementation of polar codes, and SIC. The proposed approach is further improved by randomly dividing users into different groups each of which has a unique transmitting power and an interleaving pattern. The third scheme explores a URA configuration with a receiver equipped with multiple antennas and a passive RIS, incorporating frame slotting, pilot transmission, RIS phase shift design, polar code transmission, and SIC.

In the first part of the thesis, we consider URA over GMAC, and propose a random spreading approach with polar codes, for which each user first encodes its message by a polar code, and then the coded bits are spread using a random spreading sequence. The proposed approach divides the active users into different groups, and employs different power levels for each group in such a way that the average power constraint is satisfied. We formulate and solve an optimization problem to determine the number of groups, and the corresponding expected user numbers and power levels. Extensive simulations demonstrate that the proposed

algorithm surpasses existing URA schemes in GMAC, with greater effectiveness observed particularly in scenarios with a large number of active users.

We then explore URA over Rayleigh block-fading channels with a receiver equipped with multiple antennas. We propose a slotted structure with multiple stages of orthogonal pilots, each of which is randomly picked from a codebook. In the proposed signaling structure, each user encodes its message using a polar code and appends it to the selected pilot sequences to construct its transmitted signal. Accordingly, the transmitted signal is composed of multiple orthogonal pilot parts and a polar-coded part, which is sent through a randomly selected slot. The performance of the proposed scheme is further improved by randomly dividing users into different groups each having a unique interleaver-power pair. We also apply the idea of multiple stages of orthogonal pilots to the case of a single receive antenna. In all the set-ups, we use an iterative approach for decoding the transmitted messages along with a suitable successive interference cancellation technique. The use of orthogonal pilots and the slotted structure lead to improved message recovery and reduced computational complexity in the proposed set-ups, and make the implementation with short blocklengths more viable.

In the third and final part of the thesis, we consider a URA set-up equipped with a passive RIS, where a massive number of unidentified users only a small fraction of them being active at any given time transmit their data to the BS, without any collaborations among themselves or with the BS. We introduce a slotted coding scheme, where each user chooses a slot at random for transmitting its signal. This signal consists of a pilot in its initial portion and a polar codeword that has been randomly spread in its subsequent segment. The proposed decoder operates in two phases. In the first phase, called the RIS configuration phase, the BS detects the pilots transmitted by users. The detected pilots are then utilized to estimate their corresponding CSI, using which the BS suitably selects the phase shifts of the RIS elements. In the second phase, called the data phase, transmitted messages of active users are decoded. The proposed channel estimator offers the capability to estimate the channel coefficients of the users whose pilots interfere with each other without prior access to the list of selected

pilots or the number of active users. The algorithms put forth for selecting the phase shifts of the RIS elements are specifically designed to optimize an appropriate metric within URA systems. Moreover, we extend the proposed algorithms, which are originally designed under the assumption of complete blockage of direct links between the users and the BS, to the case where direct links exist. It is illustrated that in the scenarios where the direct user-BS links are completely blocked or significantly attenuated, employing RIS improves the performance of a URA system by creating additional links between the BS and the users. The effectiveness of the proposed scheme highlights its superiority when compared to state-of-the-art RIS-aided URA schemes, demonstrating up to a 6 dB improvement in the required energy.

The studies in this dissertation can be extended in different ways. For example, the currently proposed and other existing RIS-aided URA schemes primarily focus on introducing channel coding for such scenarios. However, there remains a promising research avenue in exploring performance bounds for this configuration, which may serve as benchmarks for practical and implementable solutions.

This provides the opportunity for researchers to use these performance bounds as a benchmark for future studies.

Presently, machine learning, particularly deep learning, is finding practical use in wireless communications across tasks like resource allocation, signal processing, channel estimation, and transceiver design. Evidence suggests that the integration of machine learning can simplify the intricacies of designing wireless communication networks while achieving excellent performance. Within the context of URA, BS encounters with the challenge of handling the wireless connections of a large number of devices, leading to substantial computational complexity. The application of machine learning to URA may substantially alleviate this complexity. Therefore, a prospective avenue for future research could involve the application of machine learning based solutions to develop algorithms tailored to the URA domain.

Current URA schemes focus solely on uplink transmission. Nevertheless, incorporating feedback can enhance system performance by alerting users to decoding failures, offering them the chance for retransmission. However, the challenge lies in providing feedback to unsourced active users, as the base station lacks knowledge of their identities. Future research directions could explore novel URA schemes that integrate very limited downlink feedback to enhance the effectiveness of traditional URA approaches.

Another potential avenue for further exploration is the consideration of hardware impairments within the URA framework. To link an unbounded number of users in the URA system, it is inevitable to utilize cheap sensors, which are susceptible to impairments originating from various sources such as amplifier nonlinearities, I/Q-imbalance, phase noise, and quantization errors. Since there is only a limited amount of related work on the effect of hardware impairments on URA, exploring novel algorithms resilient to such impairments may be a practical and appealing research topic.

# Bibliography


[1] Y. Qi, A. U. Quddus, M. A. Imran, and R. Tafazolli, “Semi-persistent RRC protocol for machine-type communication devices in LTE networks,” IEEE Access, vol. 3, pp. 864–874, 2015.

[2] E. Dutkiewicz, X. Costa-Perez, I. Z. Kovacs, and M. Mueck, “Massive machine-type communications,” IEEE Network, vol. 31, no. 6, pp. 6–7, Dec. 2017.

[3] H. Shariatmadari, S. Iraji, and R. Jantti, ¨ From Machine-to-Machine Communications to Internet of Things: Enabling Communication Technologies. Chapman and Hall/CRC, 2017.

[4] N. Zhang and X. Zhu, “A Hybrid Grant NOMA Random Access for Massive MTC Service,” IEEE Internet Things J., pp. 1–1, 2022.

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

[6] J. Choi, J. Ding, N. -P. Le and Z. Ding, "Grant-Free Random Access in Machine-Type Communication: Approaches and Challenges," in IEEE Wireless Communications, vol. 29, no. 1, pp. 151-158, February 2022.

[7] G. Ma et al., "Coded Tandem Spreading Multiple Access for Massive Machine-Type Communications," in IEEE Wireless Communications, vol. 25, no. 2, pp. 75-81, April 2018.

[8] Y. Li et al., "Unsourced multiple access for 6G massive machine type communications," in China Communications, vol. 19, no. 3, pp. 70-87, March 2022.

[9] G. Miao, J. Zander, K. W. Sung, and S. B. Slimane, Fundamentals of Mobile Data Networks. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[10] H. Grama Srinath, S. Jha, and N. Mysore Balasubramanya, “An efficient NB-IoT compatible GF-NOMA PHY mechanism for mMTC,” IEEE Internet Things J., vol. 10, no. 20, pp. 17949–17963, Oct. 2023.

[11] S. Lienetal., “5G new radio: Waveform, frame structure, multiple access, and initial access,” IEEE Commun. Mag., vol. 55, no. 6, pp. 64–71, Jun. 2017.

[12] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam and S. J. Johnson, "Grant-Free Non-Orthogonal Multiple Access for IoT: A Survey," in IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1805-1838, thirdquarter 20202.

[13] G. Kasper Facenda and D. Silva, “Efficient scheduling for the massive random access Gaussian channel,” IEEE Trans. Wireless Commun., vol. 19, no. 11, pp. 7598–7609, Nov. 2020.

[14] A. T. Abusabah, N. M. Balasubramanya and R. Oliveira, "Performance Evaluation of Uplink Grant-Free Access Networks Based on Spreading-Based NOMA," in IEEE Internet of Things Journal, early access.

[15] R. Razavi, M. Dianati, et al., “Non-Orthogonal Multiple Access (NOMA) for future radio access,” 5G Mobile Communications, 2016, pp. 135–163.

[16] A. F. M. S. Shah, A. N. Qasim, M. A. Karabulut, H. Ilhan and M. B. Islam, "Survey and Performance Evaluation of Multiple Access Schemes for Next-Generation Wireless Communication Systems," in IEEE Access, vol. 9, pp. 113428-113442, 2021.

[17] Y. Chen et al., "Toward the Standardization of Non-Orthogonal Multiple Access for Next Generation Wireless Networks," in IEEE Communications Magazine, vol. 56, no. 3, pp. 19-27, March 2018.

[18] Z. Ding, R. Schober, P. Fan, and H. V. Poor, “Simple semi-grantfree transmission strategies assisted by non-orthogonal multiple access,” IEEE Trans. Commun., vol. 67, no. 6, pp. 4464–4478, Jun. 2019.

[19] N. H. Mahmood, R. Abreu, R. Böhnke, M. Schubert, G. Berardinelli, and T. H. Jacobsen, “Uplink grant-free access solutions for URLLC services in 5G new radio,” in Proc. ISWCS, Aug. 2019, pp. 607–612.

[20] C. Pyo et al., “A throughput study of grant-free multiple access for massive wireless communications,” in Proc. Int. Symp. Wireless Personal Multimedia Commun., 2017, pp. 529–534.

[21] N. Abramson, “The ALOHA system—another alternative for computer communications,” in Proc. Fall Joint Comput. Conf., Nov. 1970, pp. 281–285.

[22] L. G. Roberts, “ALOHA packet system with and without slots and capture,” SIGCOMM Comput. Commun. Rev., vol. 5, pp. 28–42, Apr. 1975.

[23] B. Zhang, Y. Zhang, Y. Wang and Z. Yang, "Value-optimal priority-aware irregular repetition slotted ALOHA in the satellite-integrated internet of things via non-cooperative game," in IEEE Internet of Things Journal.

[24] Mohamed Hadded et al. “TDMA-based MAC protocols for vehicular ad hoc networks: a survey, qualitative analysis, and open research issues”. In: IEEE Communications Surveys & Tutorials 17.4 (2015), pp. 2461–2492.

[25] 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.

[26] 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.

[27] E. Casini, R. De Gaudenzi and O. Del Rio 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.

[28] G. L. Choudhury and S. S. Rappaport, "Diversity ALOHA- A random access scheme for satellite communications," IEEE Trans. Commun., vol. 31, pp. 450–457, Mar. 1983.

[29] U. Demirhan and T. M. Duman, "Irregular repetition slotted ALOHA with energy harvesting nodes," IEEE Trans. Wireless Commun., vol. 18, pp. 45054517, September 2019.

[30] T. Akyildiz, U. Demirhan, and T. M. Duman, "Energy harvesting irregular repetition ALOHA with replica concatenation," IEEE Trans. Wireless Commun., vol. 20, no. 2, pp. 955-968, February 2021.

[31] J. Haghighat and T. M. Duman, "Analysis of coded slotted ALOHA with energy harvesting nodes for perfect and imperfect packet recovery scenarios," IEEE Trans. Wireless Commun., early access.

[32] J. Haghighat and T. M. Duman, "An energy-efficient feedback-aided irregular repetition slotted ALOHA scheme and its asymptotic performance analysis," IEEE Trans. Wireless Commun., early access.

[33] M. Mohammadkarimi, M. A. Raza, and O. A. Dobre, "Signature-based nonorthogonal massive multiple access for future wireless networks: Uplink massive connectivity for machine-type communications," IEEE Veh. Technol. Mag., vol. 13, no. 4, pp. 40–50, Dec. 2018.

[34] X. Shao, L. Cheng, X. Chen, C. Huang and D. W. K. Ng, "Reconfigurable intelligent surface-aided 6G massive access: coupled tensor modeling and sparse bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10145–10161, Dec. 2022.

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

[36] M. J. Ahmadi and T. M. Duman, "Random spreading for unsourced MAC with power diversity," in *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3995–3999, Dec. 2021.

[37] M. J. Ahmadi and T. M. Duman, "Unsourced random access with a massive MIMO receiver using multiple stages of orthogonal pilots," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Espoo, Finland, July 2022, pp. 2880-2885.

[38] M. J. Ahmadi, M. Kazemi, and T. M. Duman, "Unsourced random access with a massive MIMO receiver using multiple stages of orthogonal pilots: MIMO and single-antenna structures," *IEEE Trans. Wireless Commun.*, June 2023.

[39] M. J. Ahmadi, M. Kazemi, and T. M. Duman, "RIS-aided unsourced random access," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Kuala Lumpur, Malaysia.

[40] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[41] E. Björnson, E. G. Larsson and M. Debbah, "Massive MIMO for maximal spectral efficiency: how many users and pilots should be allocated?," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[42] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, no. 3-4, pp. 154–655, Nov. 2017.

[43] J. Hoydis, S. ten Brink and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[44] W. Yan, X. Yuan, Z. -Q. He and X. Kuai, "Passive beamforming and information transfer design for reconfigurable intelligent surfaces aided multiuser MIMO systems," IEEE J. Select. Areas Commun., vol. 38, no. 8, pp. 1793–1808, Aug. 2020.

[45] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," IEEE Trans. Wireless Commun., vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[46] A. L. Swindlehurst, G. Zhou, R. Liu, C. Pan and M. Li, "Channel estimation with reconfigurable intelligent surfaces—a general framework," *Proc. IEEE*, vol. 110, no. 9, pp. 1312–1338, Sep. 2022.

[47] X. Wei, D. Shen and L. Dai, "Channel estimation for RIS assisted wireless communications—part II: an improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, May 2021.

[48] C. Hu, L. Dai, S. Han and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7736–7747, Nov. 2021.

[49] C. Ruan, Z. Zhang, H. Jiang, J. Dang, L. Wu and H. Zhang, "Approximate message passing for channel estimation in reconfigurable intelligent surface aided MIMO multiuser systems," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5469–5481, Aug. 2022.

[50] H. Guo and V. K. N. Lau, "Uplink cascaded channel estimation for intelligent reflecting surface assisted multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 70, no. , pp. 3964–3977, Nov. 2022.

[51] X. Shi, J. Wang and J. Song, "Triple-structured compressive sensing-based channel estimation for RIS-aided MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 11095–11109, Dec. 2022.

[52] A. Glebov, P. Rybin, K. Andreev and A. Frolov, "Energy Efficiency of Unsourced Random Access Over the Binary-Input Gaussian Channel," in *IEEE Commun. Lett.*, vol. 27, no. 9, pp. 2313–2317, Sep. 2023.

[53] V. K. Amalladinne, J. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," IEEE Trans. Inf. Theory, vol. 66, no. 10, pp. 6509–6533, Oct. 2020.

[54] K. Andreev, P. Rybin and A. Frolov, "Coded Compressed Sensing With List Recoverable Codes for the Unsourced Random Access," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 7886–7898, Dec. 2022.

[55] A. K. Tanc and T. M. Duman, "Massive random access with trellisbased codes and random signatures," IEEE Commun. Lett., vol. 25, no. 5, pp. 1496–1499, May 2021.

[56] 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.

[57] V. Shyianov, F. Bellili, A. Mezghani, and E. Hossain, "Massive unsourced random access based on uncoupled compressive sensing: Another blessing of massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 820–834, Mar. 2021.

[58] 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.

[59] Z. Luan, Y. Wu, S. Liang, W. Han, B. Bai and L. Zhang, "Modulation for Massive Unsourced Random Access Based on Tensor Block Term Decomposition, " in *Proc. IEEE GLOBECOM*, pp. 637–643, Rio de Janeiro, Brazil, Jan. 2022.

[60] A. Vem, K. R. Narayanan, J.-F. 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.

[61] A. Vem, K. R. Narayanan, J. Cheng, and J.-F. Chamberland, "A user-independent serial interference cancellation based coding scheme for the unsourced random access Gaussian channel," in *Proc. IEEE ITW*, pp. 121–125, Kaohsiung, Taiwan, 2017.

[62] A. Pradhan, V. Amalladinne, A. Vem, K. R. Narayanan, and J.-F. Chamberland, "A joint graph based coding scheme for the unsourced random access Gaussian channel," in *Proc. IEEE GLOBECOM*, pp. 1–6, Waikoloa, USA, 2019.

[63] A. K. Pradhan, V. K. Amalladinne, A. Vem, K. R. Narayanan, and J. Chamberland, "Sparse IDMA: A joint graph-based coding scheme for unsourced random access," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7124–7133, Nov. 2022.

[64] 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. (ICC)*, Dublin, Ireland, June 2020, pp. 1–6.

[65] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan and J. -F. Chamberland, "LDPC codes with soft interference cancellation for uncoordinated unsourced multiple access," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Montreal, Canada, June 2021, pp. 1–6.

[66] Z. Han, X. Yuan, C. Xu, S. Jiang and X. Wang, "Sparse Kronecker-product coding for unsourced multiple access," *IEEE Wireless Commun. Lett.*, vol. 10, no. 10, pp. 2274-2278, Oct. 2021.

[67] E. Marshakov, G. Balitskiy, K. Andreev, and A. Frolov, "A polar code based unsourced random access for the Gaussian mac," in *Proc. IEEE VTC*, pp. 1–5, Honolulu, USA, 2019.

[68] K. Andreev, S. S. Kowshik, A. Frolov and Y. Polyanskiy, "Low complexity energy efficient random access scheme for the asynchronous fading MAC," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Honolulu, USA, Sep. 2019, pp. 1–5.

[69] S. S. Kowshik, K. Andreev, A. Frolov and Y. Polyanskiy, "Short-Packet Low-Power Coded Access for Massive MAC," in *Proc. 53rd Asilomar Conference on Signals, Systems, and Computers*, CA, USA, Sep. 2019, pp. 827–832.

[70] X. Chen, D. Guo, and G. W. Wornell, "Sparse OFDM:A Compressive Sensing Approach to Asynchronous Neighbor Discovery," arXiv preprint 1706.09387, 2017.

[71] V. K. Amalladinne, K. R. Narayanan, J. -F. Chamberland and D. Guo, “Asynchronous Neighbor Discovery Using Coupled Compressive Sensing,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May. 2019, pp. 4569–4573.

[72] A. Fengler, O. Musa, P. Jung, and G. Caire, ”Pilot-based unsourced random access with a massive MIMO receiver, interference cancellation, and power control,” *IEEE J. Sel. Areas Commun.*, vol. 40, no. 5, pp. 1522–1534, May 2022.

[73] M. Gkagkos, K. R. Narayanan, J. F. Chamberland and C. N. Georghiades, ”FASURA: A Scheme for Quasi-Static Fading Unsourced Random Access Channels,” *IEEE Trans. Commun.*, July 2023.

[74] D. Ustinova, A. Frolov and K. Andreev, “Unsourced Random Access Pilot-Assisted Polar Code Construction for MIMO Channel, ” in *Proc. IEEE SIBIRCON*, pp. 637–643, Yekaterinburg, Russian Federation, Nov. 2022.

[75] Y. Su and J. Zheng, ” Index Modulation Multiple Access for Unourced Random Access,” *IEEE Wireless Commun. Lett.*, vol. 12, no. 5, pp. 794–798, May 2023.

[76] Y. Su, J. Zheng, Z. Liang and J. Ni, ” Massive MIMO Enabld Unsourced Random Access Through LDPC Coding and Random Spreading," in *Proc. IEEE ICCT*, pp. 178–183, Nanjing, China, Nov. 2022.

[77] M. Ozates, M. Kazemi and T. M. Duman, ”A Slotted Pilot-Based Unsourced Random Access Scheme with a Multiple-Antenna Receiver,” *IEEE Trans. Wireless Commun.*, Aug. 2023.

[78] Z. Han, X. Yuan, C. Xu and X. Wang, “Receiver design for MIMO unsourced random access with SKP coding,” *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 45-49, Jan. 2023.

[79] J. C. Jiang and H. M. Wang, “A Fully Bayesian Approach for Massive MIMO Unsourced Random Access,” *IEEE Trans. Commun.*, vol. 71, no. 8, pp. 4620–4635, Aug. 2023.

[80] K. Andreev, E. Marshakov and A. Frolov, "A polar code based TIN-SIC scheme for the unsourced random access in the quasi-static fading MAC," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, USA, June 2020, pp. 3019–3024.

[81] S. S. Kowshik and Y. Polyanskiy, "Quasi-static fading MAC with many users and finite payload," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, July 2019, pp. 440–444.

[82] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user MAC with finite payloads and fading," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 5853–5884, Sep. 2021.

[83] M. Kazemi, T. M. Duman, and M. Médard, "Collision resolution for random access," IEEE Trans. Wireless Commun., vol. 21, no. 5, pp. 3464–3477, May 2022.

[84] M. Kazemi, T. M. Duman, and M. Medard, "Double-zipper: Multiple access with ZigZag decoding," in Proc. IEEE Int. Conf. Commun. (ICC), Dublin, Ireland, Jun. 2020, pp. 1–6.

[85] S. S. Kowshik, K. Andreev, A. Frolov and Y. Polyanskiy, "Energy efficient coded random access for the wireless uplink," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4694–4708, Aug. 2020.

[86] 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 (ISIT)*, Paris, France, July 2019, pp. 2768–2772.

[87] M. Ozates and T. M. Duman, "Unsourced random access over frequency-selective channels," *IEEE Commun. Lett.*, vol. 27, no. 4, pp. 1230–1234, Apr. 2023.

[88] J. Gao, Y. Wu, T. Li and W. Zhang, "Energy Efficiency of MIMO Massive Unsourced Random Access With Finite Blocklength," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 743-747, April 2023.

[89] M. Zheng, Y. Wu, and W. Zhang, "Polar coding and sparse spreading for massive unsourced random access," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Victoria, Canada, Feb. 2020, pp. 1–5.

[90] Y. Polyanskiy, H. V. Poor and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[91] E. Arikan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, May 2009.

[92] H. Shen, W. Xu, S. Gong, Z. He and C. Zhao, "Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1488–1492, Sept. 2019.

[93] F. Laue, V. Jamali and R. Schober, "IRS-assisted active device detection," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Lucca, Italy, Sep. 2021, pp. 556–560.

[94] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, June 2019.

[95] X. Wei, D. Shen and L. Dai, "Channel estimation for RIS assisted wireless communications—part II: an improved solution based on double-structured sparsity," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1403–1407, May 2021.

[96] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase-shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, March 2020.

[97] Q. Wu and R. Zhang, "Intelligent Reflecting Surface Enhanced Wireless Network: Joint Active and Passive Beamforming Design," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[98] P. D. Lax, Linear Algebra and Its Applications. Hoboken, NJ, USA: Wiley-Interscience, 2007.

[99] K. M. Hoffman and R. Kunze, Linear algebra. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.

[100] W. E. Roth, "On direct product matrices," Bulletin of the American Mathematical Society, vol. 40, pp. 461–468, 1934.

[101] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.