SAPIENZA
UNIVERSITÀ DI ROMA

DIPARTIMENTO DI
INGEGNERIA DELL'INFORMAZIONE
ELETTRONICA E TELECOMUNICAZIONI

# Code-domain Non Orthogonal Multiple Access for 5G networks


By

Mai Thi-Phuong Le

Sapienza University of Rome

Department of Information Engineering, Electronics and Telecommunications


Supervisor

Prof. Maria-Gabriella Di Benedetto

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Telecommunications

October 2018

**Code-domain Non Orthogonal Multiple Access for 5G networks**
Ph.D. thesis. Sapienza – University of Rome



This thesis has been typeset by LaTeX and the LaTeX $2_\varepsilon$ class `sapthesis.cls` developed by Francesco Biccari with the help of GuIT (Gruppo Utilizzatori Italiani di TeX).

Version: January 22, 2019

Author's email: mai.le.it@ieee.org, pmai2810@gmail.com

*Dedicated to*
*little Chip, Nam and my parents*

# Acknowledgments

*As a friend of mine once said: Doing a PhD is a journey full of ups and downs. For me, PhD is not only a journey full of emotions but will definitely be a wonderful experience of my life. To arrive this point of the PhD journey, I felt myself extremely lucky to have met with many people that I will may achieve nothing without them. I would like to thank everyone who supported me on this journey, especially to those whom I unintentionally do not remember at the time of writing.*

*Foremost, I would like to express my deepest gratitude to my supervisor Prof. Maria-Gabriella Di Benedetto for her guidance and support both literally and mentally. I was fully inspired with her working attitude, great intuition, and strong determination. Those extraordinary qualities have shaped my thought that may benefit me on the way to become an independent researcher. It will remain a blessed memory about the time we were sitting for hours in her office while writing papers or discussing ideas or life.*

*I would like to thank Luca for his kindness. He has been always helpful to me, especially the first period of my arrival in Rome. I enjoyed very much the lunch-time discussions with him, who is always a funny and optimistic person in me.*

*I sincerely thank my dear lab-mate Giuseppe for accompanying me during this journey. I feel so lucky to have him as a friend, a collaborator, and even an Italian-English translator, who always arises when I need help. I will always remember our endless discussions during coffee break about papers, academic career and life. The PhD life would be so boring without such a great friend. My thanks also extend to my new lab-mate Alireza, the department staffs Nicola and Anna Paola, for being nice to me.*

*My special thank goes to Guido, who treated me so well during my research visiting in Singapore University of Technology and Design (SUTD). He is not just a dear friend, but a mentor, who helped to build first steps in my early PhD stage. Two months working with Guido is not a long time, but what I learned from him is precious: preciseness in each claim/sentence, critical thinking, long-term plans, that are all of necessity for being a researcher.*

*My gratitude extends to Prof. Tony Q. S. Quek for welcoming me to his research group, making my stay in SUTD a pleasant and rewarding experience. I was impressed by his immense intuition in many broad fields of wireless communications, his work productivity and his kindness. I also want to thank my recent friends in Tony's group: Duong, Khai, Thinh, Mao, Chan, Dzung and others for being such nice friends and for making my Singapore memory so interesting. Their dedicated attitude to work will be a sample for me to learn from.*

*I am grateful to my subcommittee consisting of Prof. Vincenzo Eramo, Prof. Antonio Cianfrani and Prof. Massimo Panella, who have been followed my PhD activities since the first year, and have gave many constructive advices during annual meetings. I would like to take this chance to thank the two external reviewers of my thesis for their valuable time and their possible feedback.*

*My sincere thanks are dedicated to my colleagues in Department of Electronic and Telecommunications, Danang University of Science and Technology, Vietnam for their continuous support during my PhD program, especially to Prof. Nguyen Van Tuan, the dean of the department.*

*I want to thank my best friends Tu, who always makes me delightful by her own subtle way and Nghia, who is always close to me despite the distance. My thanks expand to my dear friends Duong, Binh, and particularly to Thai for being a great soul-mate.*

*Many thank is given to my older sister Ha, whom I know I can lean on anytime. I could have overcome the most difficult moment of the last year PhD due to family situation since I am sure that she can take care of my parents while I am far away from home. My gratitude is also sent to my little sister Nhi, who both physically and mentally took care of my son while I was traveling.*

*This thesis cannot come out without the strong effort of my beloved Mom. Her will and the persistence of my Dad indeed encouraged me to go through the last year of PhD.*

*Last but not least, I am so grateful to my dear husband, Nam, who just finished PhD one year before me, for sharing everything: knowledge, experience of doing PhD, housework, bringing our son up, and tears, and laughs. Having you and Chip beside me in Rome is the luckiest thing of the PhD life. Of course, this thesis is dedicated to our endless love: little Chip. He is the most precious gift that I would thank God forever to bring me such full of joy and happiness. I hope that someday this thesis will be the starting point to bring him to the beautiful world: the world of wisdom.*

# Contents

# Synopsis

While expected to be standardized by the year 2020, the fifth generation (5G) currently receives considerable attention from the wireless community [1]. Among the key features chacracterizing 5G, non-orthogonal multiple access (NOMA) is one of the promising technologies, that are expected to address the targets of 5G wireless communications, including high spectral efficiency, massive connectivity, and low latency [2, 3].

Back to the history of cellular communications from 1G to 4G, the radio multiple access schemes are mostly characterized by orthogonal multiple access (OMA), where different users are assigned to orthogonal resources in either frequency (frequency-division multiple access (FDMA) and orthorgonal FDMA (OFDMA)), time (time-division multiple access (TDMA)) or code (synchronous code-division multiple access (CDMA) in underloaded condition) domains. However, 5G multiple access is required to support a wide range of use cases, providing access to massive numbers of low-power internet-of-thing (IoT), as well as broadband user terminals in the cellular network. Providing high spectral efficiency, while minimizing signaling and control overhead to improve efficiency, may not be feasible to achieve by OMA techniques [4]. In fact, the orthogonality condition can be imposed as a requirement only when the system is *underloaded*, that is, when the number of active users is lower than the number of available resource elements (degrees of freedom or dimensions).

The idea of NOMA is to serve multiple users in the same band and abandon any attempt to provide orthogonal access to different users as in conventional OMA. Orthogonality naturally drops when the number of active users is higher than the number of degrees of freedom, and "collisions" appear. One possible way of controlling collisions in NOMA is to share the same signal dimension among users and exploit power (power-domain NOMA (PDM-NOMA)) vs. code (code-domain NOMA (CDM-NOMA)) domains [2]. However, refer to NOMA, most of intuition gained from the recent literature implies power-domain case [5], which was firstly introduced by Mazzini [6], including integration of NOMA with other technologies such as MIMO-NOMA, Cognitive Radio NOMA (CR-NOMA), mm-Wave NOMA, full-duplex NOMA and so on.

In PDM-NOMA, it uses superposition coding, a well-known non-orthogonal scheme for downlink transmissions [7], and makes superposition decoding possible by allocating different levels of power to different users [8]. The "near" user, with a higher channel gain, is typically assigned with less transmission power, which helps making successive interference cancellation (SIC) affordable at this user [9]. Interested readers are referred to the latest works on PDM-NOMA such as [5, 10].

CDM-NOMA is characterized by sparsity employed in spreading sequences or multi-dimensional codewords. It is worthy noting that CDM-NOMA and conventional CDMA share the same working principle in exploiting different spreading codes. As a matter of fact, several characterizing variants of CDM-NOMA, such as low-density spreading CDMA (LDS-CDMA) [11–13], low-density spreading orthogonal frequency-division multiplexing (LDS-OFDM) [14], sparse code multiple access (SCMA) [15], pattern division multiple access (PDMA) [16], and multi-user shared access (MUSA) [17], may be inferred from the framework of CDMA. By relaxing orthogonality requirements, CDM-NOMA variants enable flexible resource allocation, and reduce hardware complexity.

This thesis aims to shed some light on understanding CDM-NOMA and its different dialects, particularly the schemes with single-carrier waveforms from an information-theoretic perspective. At the moment, NOMA has been currently proposed for the 3rd generation partnership project long-term evolution advanced (3GPP-LTE-A) standard, the next general digital TV standard (ATSC 3.0), and the 5G New Radio (NR) standard. In fact, CDM-NOMA variants are currently under consideration via Specification TS 38.812 (Study on NOMA for NR) in anticipation to have a "ready" NR system in 2020 [18]. The emergence of a complete theoretical work on CDM-NOMA is, therefore, of essence and of expectation to contribute as a timely reference for future release of 5G standardization.

In order to understand CDM-NOMA[1] in terms of fundamental limits, the considered framework focuses on investigating the following issues:

- the impact of system load, which classifies NOMA vs. OMA,
- the impact of sparsity, which classifies NOMA further into low-dense vs. dense,
- the impact of regularity, which characterizes possible spreading mapping constraints,
- the impact of the channel fading, especially flat-fading, which is very common in practical scenarios,
- the impact of the channel knowledge, known as channel state information (CSI), which is characterized by the rapid change of real-world communication channels.

The first three issues are investigated subject to the ideal assumption of AWGN channel, and the rests are studied subject to the flat-fading channel assumption, respectively.

The thesis starts with a mathematical framework (Chapter 1), which is proposed based on the seminal framework for the conventional CDMA including direct-sequence CDMA (DS-CDMA) and multi-carrier CDMA. Based on this framework, information-theoretic results are expected to explore the relationship between the achievable rates and the aforementioned peculiar features of NOMA.

Chapter 2 divides different NOMA variants into single-carrier and multi-carrier systems, corresponding to DS-CDMA and MC-CDMA, and further classifies NOMA

[1] Below CDM-NOMA is called shortly as NOMA for simplicity

schemes as low-dense vs. dense groups due to sparsity of the representation matrix of each scheme. The representation matrix $\boldsymbol{S}$, well-known as spreading matrix in DS-CDMA, is considered as the fundamental element of the proposed framework since it defines different NOMA variants. Each NOMA variant is then described in detailed and shown to be a close relative with the traditional CDMA.

The next chapters (Chapter 3–5) focus on single-carrier NOMA schemes, particularly on theoretical behavior of low-dense system. Based on the proposed framework, theoretical analysis of NOMA is investigated considering massive connectivity of 5G, i.e. the number of users $K$ is supposed to be very large compared to the number of REs $N$. The behavior of the system should thus be considered in the asymptotic limit, where both $K$ and $N$ go to infinity, while the ratio $K/N = \beta$, called load, remains finite. This corresponds to analyzing the system in the *large system limit* (LSL) [19].

In Chapter 3, theoretical behavior of three distinctive cases, including dense vs. so-called *irregular* and *regular* low-dense NOMA in the LSL, corresponding to DS-CDMA ($N_s = N$) [20] vs. LDS-CDMA ($N_s = 1$) [21] and *regular* sparse NOMA ($2 \leq N_s \in \mathbb{N}^+$) [22], is investigated for the AWGN channels via existing closed-form expressions, with respect to load, sparsity and regularity. It will be shown that in AWGN channels, *irregular* low-dense NOMA, for example LDS-CDMA, is superior to all the rest in terms of spectral efficiency, given the suboptimal receivers (linear receivers in this case) are adopted.

In the follow-up Chapter 4, *irregular* low-dense NOMA, as a spectral-efficient representative, is investigated under the impact of flat-fading with perfect knowledge of channel (perfect CSI) for both optimum and linear receivers. On the other hand, Chapter 5 addresses the impact of a practical issue of no prior knowledge of propagation channel, i.e. channel without CSI. In this chapter, capacity bounds of low-dense NOMA are provided in the context of 5G New Radio, considering the impact of number of coherence symbols, number of users, system load and SNR. Finally, Chapter 6 concludes the thesis and discusses the possible future works.

Main contributions of the thesis can be listed as:

- A unified framework for several code-domain NOMA schemes is presented. The analytical framework, built on the traditional CDMA model, proved to be flexible enough for representing several code-domain dialects, and, in particular, addressed properties of a fundamental element of the model, that is, the representation matrix $\boldsymbol{S}$. From this proposed model, the theoretical analysis is investigated for both optimum and linear receivers in the LSL, taking account of the impact of aforementioned issues.

- For AWGN channels, the comparative analysis will show that it is beneficial to adopt extreme low-dense single-carrier NOMA in the LSL. Particularly, when optimum receivers are used, the adoption of a *regular* low-dense spreading matrix is worthwhile to the system achievable rates, which are higher than those obtained with either *irregular* low-dense or dense formats, for any value of load. For linear receivers, which are more favorable in practice due to low complexity, the *regular* low-dense NOMA still has better performance in the

underloaded regime (load < 1), while the *irregular* counterpart outperforms all the other schemes in the overloaded scenario (load > 1).

- For flat-fading channels (for e.g. Rayleigh fading) with perfect CSI, the behavior of a typical single-carrier NOMA, that is, low density spreading NOMA (LDS) is compared against dense spreading (DS) one. In the presence of flat-fading, [19] showed that in the overloaded condition, the fading effect, for DS, is to enhance spectral efficiency. What missing is the behavior of low-dense system under the impact of fading. We filled the gap by showing that, in the LSL, irregular LDS with fading achieved higher spectral efficiency than DS in the overloaded mode subject to linear receivers. Moreover, the dominance of irregular LDS-CDMA over DS in terms of spectral efficiency performance than DS goes stronger when $E_b/N_0$ increases.
- For Rayleigh block-fading channels without CSI, information-theoretical bounds of LDS are derived. The capacity upper bound, defined as the capacity of low-dense NOMA with perfect CSI, is found from Chapter 4. The capacity lower bound is derived using a pilot-based communication scheme, as suggested in [23]. Upper and lower capacity bounds are described as a function of $E_b/N_0$, number of coherent symbols $n_b$, and system load $\beta$. The effect of the number of users $K$ is also investigated. Results indicate that, when the above factors are favorably combined, the gap between the upper and lower capacity bounds becomes negligible. In particular, when the number of coherent symbols $n_b$ is high, while the number of users simultaneously joining the network is low, the lower capacity bound well approximates the capacity with perfect CSI, leading to the conclusion that the system is still robust despite the absence of knowledge on the channel.

The results of this thesis have been partly published in the following articles:

Chapter 1–3:

- **Mai T. P. Le**, Guido Carlo Ferrante, Giuseppe Caso, Luca De Nardis, Maria-Gabriella Di Benedetto, "On Information-theoretic limits of Code-domain NOMA for 5G," IET Commun., 2018, Vol. 12 Iss. 15, pp. 1864–1871.

Chapter 4:

- **Mai T. P. Le**, Guido Carlo Ferrante, Tony Q.S. Quek, Maria-Gabriella Di Benedetto, "Fundamental Limits of Low-Density Spreading NOMA with Fading," IEEE Trans. on Wireless Commun., 2018, 17, (7), pp. 4648–4659.

Chapter 5:

- **M. T. P. Le**, G. Caso, L. De Nardis, A. Mohammadpour, G. Tucciarone, M.-G. Di Benedetto, "Capacity bounds of Low-Dense NOMA over Rayleigh fading channels without CSI," in IEEE Proc. Int. Conf. on Telecom. (ICT), pp. 428–432, St. Malo, France, 26–28 June 2018.

# Chapter 1

# A mathematical model of Code-domain NOMA

## 1.1 Introduction

Apart from the fact non-orthogonality feature of NOMA has been used recently [24], initial NOMA concept has a long story from the beginning of 1990s. Non orthogonal signal sets with particular structures were invented such that they may be detectable at the receiver. First investigations on designing guidelines for non-orthogonal spreading codes were made by Ross and Taylor [25, 26] that were applicable to an overloading system. By adding additional linearly dependent codes while maintaining the orthogonal minimum distance (Euclidean distance) to ensure feasible detection, the signal sets essentially became non-orthogonal. In addition, these sets were put under the constraint as such all users should not have higher powers than that of orthogonal set. On the receiver side, iterative decoding algorithm such as message passing algorithm (MPA) detector, also known as belief propagation (BP) algorithm in low-density parity-check (LDPC) codes, is employed.

CDM-NOMA exploits code to distinguish different users at the receiver, i.e. it function similarly to the traditional DS-CDMA system. The main feature identifying CDM-NOMA from CDMA is by employing sparsity in spreading sequences via low-density or low cross-correlation sequences. Based on the specific scenario, single-carrier or multi-carrier NOMA schemes can be adopted, corresponding to single-carrier DS-CDMA or multi-carrier CDMA (MC-CDMA), respectively. The 15 existing proposals of NOMA proposed for the Rel-14 3GPP NR Study item [27] (Table 1.1) and recent CDM-NOMA proposals available in the literature, therefore, will be classified based on dense vs. low-dense and single vs. multi carrier features.

In addition to proposed methods for NOMA from Table 1.1, including MUSA [17], PDMA [16, 28], IGMA [29], IDMA [30, 31], other NOMA schemes such as LDS-CDMA [11] (also known as time-hopping (TH-CDMA) [21, 32]), SAMA [33] will be shown tightly correlated to traditional single-carrier DS-CDMA, therefore, are classified as single-carrier NOMA. By contrast, LDS-OFDM [34, 35], along with remained NOMA schemes in Table 1.1 including SCMA [15, 36], LSSA [37], NCMA [38], NOCA [39], GOCA [40], LDS-SVE [41], LCRS/FDS [42], RDMA [40] are well fitted to multi-carrier CDMA model, hence, classified here as multi-carrier NOMA, as will be shown in

| | NOMA schemes | Full Name | Company | UL/DL |
|---|---|---|---|---|
| 1 | Power-domain NOMA | Power-domain non-orthogonal multiple access | DCM | UL/DL |
| 2 | SCMA | Sparse code multiple access | Huawei | UL/DL |
| 3 | MUSA | Multi-user shared access | ZTE | UL/DL |
| 4 | PDMA | Pattern division multiple access | CATT | UL/DL |
| 5 | LSSA | Low code rate and signature based shared access | ETRI | UL |
| 6 | RSMA | Resource spread multiple access | Qualcomm | UL |
| 7 | IGMA | Interleave-grid multiple access | Samsung | UL/DL |
| 8 | IDMA | Interleave division multiple access | Nokia | UL |
| 9 | NCMA | Non-orthogonal coded multiple access | LGE | UL |
| 10 | NOCA | Non-orthogonal coded access | Nokia | UL |
| 11 | GOCA | Group orthogonal coded access | MTK | UL |
| 12 | LDS-SVE | Low density spreading - signature vector extension | Fujitsu | UL/DL |
| 13 | FDS | Frequency domain spreading | Intel | UL |
| 14 | LCRS | Low code rate spreading | Intel | UL |
| 15 | RDMA | Repetition division multiple access | MTK | UL |

Table 1.1. NOMA schemes proposed for the Rel-14 3GPP NR Study Item [27]

Chapter 2. RSMA [43, 44] is a NOMA dialect that is proposed for both type of waveforms, depending on the specific application scenario.

## 1.2 A reference mathematical model

### 1.2.1 Single-carrier NOMA

For single-carrier NOMA, each data symbol of user $k$ is spread by $N$ chips of the corresponding spreading sequence, that is similar to as DS-CDMA system. Naturally, the mathematical model of single-carrier NOMA may be built from the baseline model of DS-CDMA, proposed by Verdú and Shamai in [19] and [20] as follows

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{H}\boldsymbol{b} + \boldsymbol{n}, \tag{1.1}$$

where the received signal $\boldsymbol{y} \in \mathbb{C}^N$ belongs to a space characterized by $N$ signal dimensions. Note that $N$ also represents the number of elements over which each symbol is spread, that is the number of REs, and equivalent to the number of

'chips', as termed commonly in CDMA. Vector $\boldsymbol{b} = [b_1, \ldots, b_K]^T \in \mathbb{C}^K$ is the vector of symbols transmitted by $K$ users. Being a random spreading matrix, $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K] \in \mathbb{C}^{N \times K}$ is composed of $K$ columns, each being the spreading sequence $\mathbf{s}_k$ of user $k$ ($1 \leqslant k \leqslant K$). Supposing the channel is flat, the channel matrix can be represented as $\boldsymbol{H} = \text{diag}\,[h_1, \ldots, h_K] \in \mathbb{C}^{K \times K}$, whereas it reduces to the identity matrix if the AWGN channel is assumed. Lastly, the noise $\boldsymbol{n} \in \mathbb{C}^N$ is described by a circularly-symmetric Gaussian vector with zero mean and covariance $N_0\boldsymbol{I}$.

The nature of the representation matrix $\boldsymbol{S}$ defines the multiple access methods. This matrix is known as spreading matrix or signature matrix in DS-CDMA, TH-CDMA, LDS-CDMA, SAMA, or code matrix in MUSA, and pattern matrix in PDMA. As a matter of fact, NOMA schemes can be classified into dense vs. low-dense, where the corresponding matrix $\boldsymbol{S}$ is dense if all REs are used vs. low-dense, when some REs are not used. In terms of energy, this corresponds to having all REs contain signal energy vs. energy is concentrated on only part of the available REs, reflected by the presence of nonzero entries in $\boldsymbol{S}$. According to this understanding, DS-CDMA inherently stands for dense spreading. For single-carrier NOMA, the dense group includes DS-CDMA and IDMA, while the low-dense group includes LDS-CDMA, TH-CDMA, SAMA, MUSA, IGMA and PDMA.

### 1.2.2 Multi-carrier NOMA

For multi-carrier NOMA schemes, the system can be described as a combination of CDMA for dense NOMA (respective, LDS-CDMA for low-dense NOMA) with multi-carrier modulation, for e.g. OFDM, that makes multi-carrier NOMA work analogously to MC-CDMA [45]. Based on this intuition, the system model of multi-carrier NOMA basically can be elaborated as below, assuming the number of OFDM subcarriers shared by every user is also equal to the spreading gain $N$ [46].
In multi-carrier NOMA, each data symbol of user $k$ is replicated into $N$ parallel copies, each copy is then multiplied by a chip from the respective spreading sequence. All $N$ copies are then mapped to $N$ subcarriers and are transmitted in parallel. Thanks to inverse discrete Fourrier transform (IDFT) implementation, those $N$ parallel chips are converted into serial sequence for further transmission [46].

Adopting the same notation of spreading matrix $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K]$ with $\boldsymbol{s}_k$ being frequency-domain spreading sequence of user $k$ as in eq. (1.1), the baseband signal by the $k$th user in time-domain is expressed by

$$\boldsymbol{W}\boldsymbol{s}_k b_k.$$

Here $\boldsymbol{W}$ denotes the $N \times N$ IDFT matrix, and $b_k$ again stands for the data symbol of the $k$th user.

As a matter of fact, the receiver consists of $N$ matched filters, corresponding to $N$ subcarriers, is equivalently to conducting a discrete Fourrier transform (DFT) on the discrete baseband domain. Therefore, the received vector by user $k$ in the frequency domain is $b_k\tilde{\boldsymbol{s}}_k$, where

$$\tilde{\boldsymbol{s}}_k = \text{diag}[h_1^1, \ldots, h_k^N]\boldsymbol{s}_k,$$

with $h_k^i$ being the fading coefficient at subcarrier $i$ of the $k$th user.

Since in practice, each subcarrier is narrow enough to experience only flat fading, the system model of multi-carrier NOMA is, therefore, mathematically equivalent to that of single-carrier NOMA in (1.1) with respect to flat-frequency fading. This observation is also reported by Tulino *et al.* [46] for the case of DS-CDMA and MC-CDMA. Here the spreading matrix $\boldsymbol{S} \in \mathbb{C}^{N \times K}$ again defines different multi-carrier NOMA schemes.

# Chapter 2

# Code-domain NOMA classification

In this chapter, code-domain NOMA schemes available at the time of writing are described. Due to the adopted waveform, NOMA schemes are classified as single- vs. multi-carrier system, each group is further divided into dense vs. low-dense based on the property of the representation matrix as discussed in Chapter 1. Code-domain NOMA classification is illustrated in Fig. 2.1.

## 2.1 Single-carrier NOMA (SC-NOMA)

Different variants of single-carrier NOMA are shown below with their corresponding example of representation matrix $\boldsymbol{S}$ as in Figure 2.2.

### 2.1.1 Dense SC-NOMA

#### Direct-sequence CDMA (DS-CDMA)

In DS-CDMA, spreading codes spread out the energy of signals over all the $N$ available REs [20]. Each column $\boldsymbol{s}_k$ of the representation matrix $\boldsymbol{S}$ is formed by a spreading code, corresponding to a single user $k$ over $K$ users. Two typical examples representing for DS-CDMA spreading codes were described by Verdú and Shamai [20], including binary and spherical sequences. In the binary sequence model, $\boldsymbol{s}_k$ is uniformly filled with $N$ values belonging to the set $\{+1/\sqrt{N}, -1/\sqrt{N}\}$, whereas $\boldsymbol{s}_k$ is modeled as a unitarily invariant unit-norm vector in the spherical model [20]. Matrix $\boldsymbol{S}$ has the dense structure, as illustrated in Fig. 2.2, for the binary case.

#### Interleave Division Multiple Access (IDMA)

IDMA was shown to be a special case of the traditional CDMA employing a common spreading code and a user-specific random interleaver [30,31]. In particular, each DS-CDMA spreading code is replaced by a length-$N$ spreading code followed by a chip-interleaver, where the interleavers should be generated independently, randomly, and uniquely for different users [30]. The system model of IDMA can be described

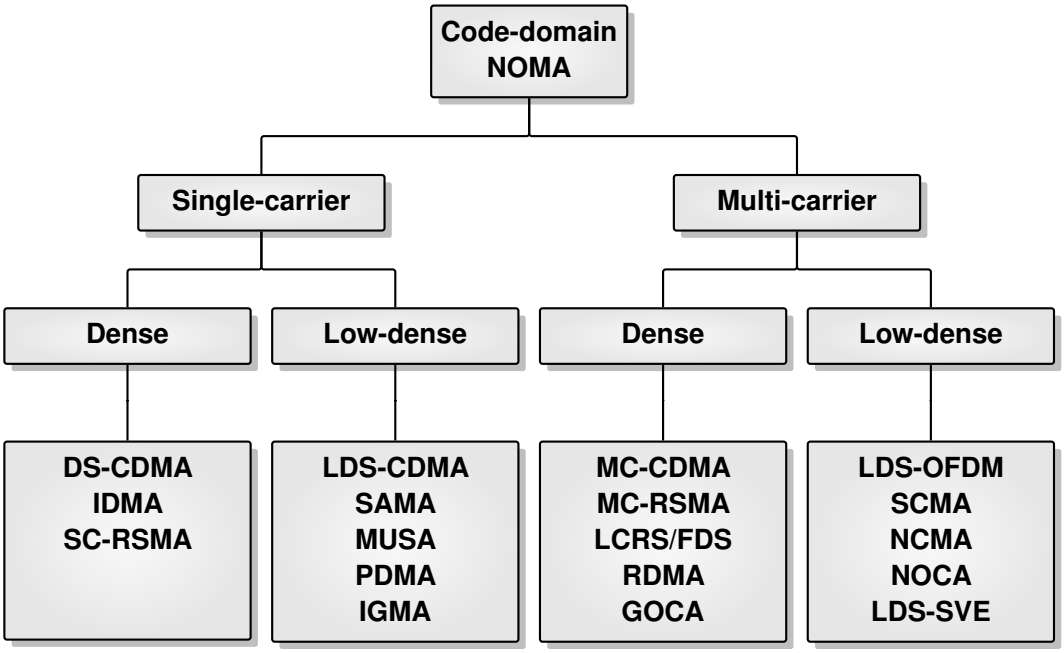



Figure 2.1. Code-domain NOMA classfification

comparably to CDMA as follows. In CDMA, the information bit $b_k$ of the user $k$-th is spread by the corresponding spreading code $\boldsymbol{s}_k$, and the users are separated easily by their respective codes. On the other hand, the user bit $b_k$ of IDMA transmitter is first spread by the *same* spreading code $\boldsymbol{s} = \boldsymbol{s}_k, \forall k$, and then permuted by a user-distinct interleaver $\pi_k$. This scheme relies on interleaving process for user separation, thus, is called as interleave division multiple access. The descriptive transmitter model of IDMA is presented in comparison with that of CDMA in Fig. 2.3.

By sharing a single spreading code for all users, the receiver complexity of IDMA was shown to be lower than CDMA, while higher spectral efficiency can be achieved [30, 31]. With optimal power allocation, the IDMA system capacity may reach the maximum capacity of the multiple access CDMA for binary AWGN channel, given the same input constraints [30]. By adopting the similar strategy of separating users, the representation matrix $\boldsymbol{S}$ of IDMA is shown identically to that of DS-CDMA (see Fig.2.2).

### Single-carrier Resource Spread Multiple Access (SC-RSMA)

RSMA is another term of NOMA, where unique signatures attached to users may be detected at the receivers by power, spreading or scrambling codes, interleavers or their combinations [43]. In [44], RSMA employs low rate channel code combining with scrambling code (by interleavers) as in IDMA to separate different users. Based on particular scenario, RSMA can be adopted as a single or multi-carrier system (see Section 2.2). For SC-RSMA, the scheme is favorable for optimized battery power consumption and link budget extension by using single carrier waveforms and low peak to average power ratio (PAPR) modulations. Furthermore, SC-RSMA is a potential solution for uplink transmission since it allows asynchronous access

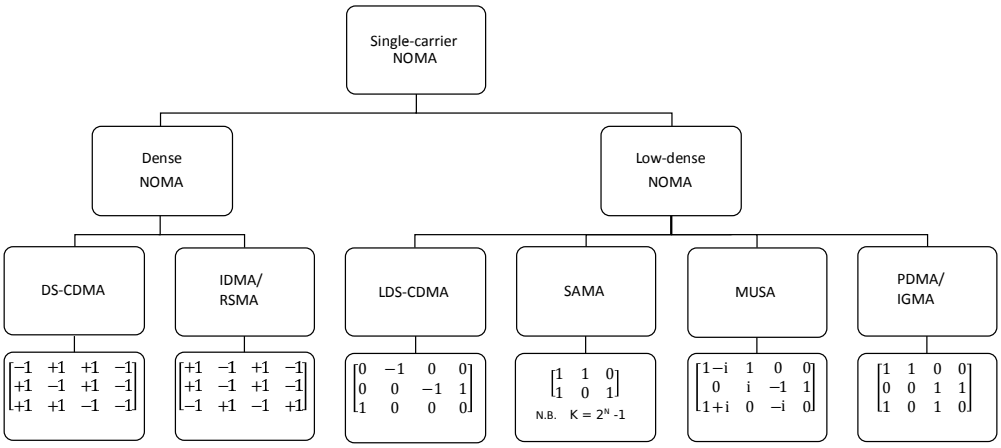



Figure 2.2. Single-carrier NOMA classification with a corresponding example of matrix for each NOMA dialect

and grant-free transmission.

### 2.1.2 Low-dense SC-NOMA

#### Low-Density Spreading CDMA (LDS-CDMA) and Time-Hopping CDMA (TH-CDMA)

LDS-CDMA replaces dense spreading sequences of DS-CDMA by sparse counterparts, as such the dense matrix $\boldsymbol{S}$ in DS-CDMA becomes sparse in LDS case [11]. Consequently, the energy of LDS-CDMA signals concentrates on a part of the REs, in lieu of uniformly spreading over all REs, as in DS-CDMA. The idea on exploiting the sparsity of multiple access based on code domain came from sparse CDMA, which was firstly investigated via statistical physics in [47] and [48], and subsequently studied in [49, 50].

It is worthy to note that, the same sparse structure can be observed in the time-hopping CDMA model [21], hence, both can be depicted by a matrix $\boldsymbol{S}$, composed by $N$ REs with only $N_s \ll N$ nonzero elements in the set $\{+1/\sqrt{N_s}, -1/\sqrt{N_s}\}$, while the remaining $(N - N_s)$ REs are zeros (see Fig. 2.2[1]).

#### Successive Iterative Cancellation Amenable Multiple Access (SAMA)

SAMA is a special case of LDS-CDMA [33]. Here, $K$ is fixed to $K = 2^N - 1$, that is, the system is necessarily overloaded. In this scheme, the nonzero element of SAMA spreading sequences are equal to one, and the matrix $\boldsymbol{S}$ should meet the two following rules [33]:

- the size of the groups having different number of 1's in the spreading sequence should be maximized,

[1] In Fig. 2.2, LDS-CDMA also stands for TH-CDMA

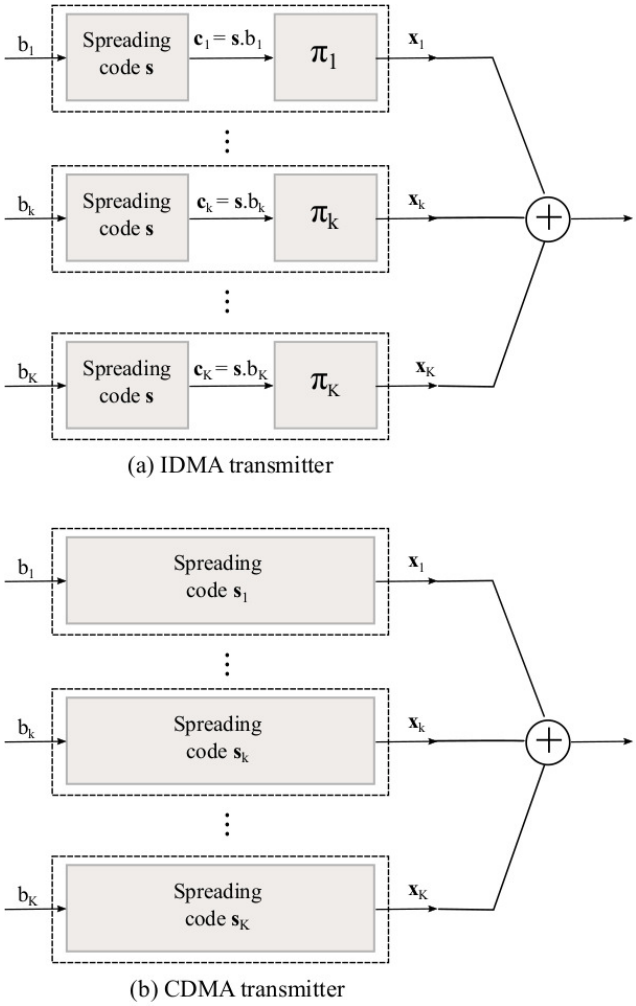



Figure 2.3. IDMA vs. CDMA transmitter

- the number of overlapped spreading sequences having the same number of 1's should be minimized,

from which the maximum number of users $K$ is computed as

$$\binom{N}{1} + \binom{N}{2} + \cdots + \binom{N}{N} = 2^N - 1.$$

Example of matrix $\boldsymbol{S}$ with $N = 4, K = 15$ for SAMA case is:

$$\boldsymbol{S}_{\text{SAMA}}^{4\times 15} = \begin{bmatrix} \underbrace{1}_{\binom{4}{4}=1} & \underbrace{1\ 0\ 1\ 1}_{\binom{4}{3}=4} & \underbrace{1\ 1\ 0\ 0\ 1\ 0}_{\binom{4}{2}=6} & \underbrace{0\ 0\ 0\ 1}_{\binom{4}{1}=4} \\ 1 & 1\ 1\ 0\ 1 & 1\ 0\ 0\ 1\ 1\ 0 & 0\ 1\ 0\ 0 \\ 1 & 1\ 1\ 1\ 0 & 0\ 1\ 0\ 1\ 0\ 1 & 0\ 0\ 1\ 0 \\ 1 & 0\ 1\ 1\ 1 & 0\ 0\ 1\ 0\ 1\ 1 & 0\ 0\ 0\ 1 \end{bmatrix}.$$

Again matrix $\boldsymbol{S}$ contains '0' and '1' elements, with a specific structure designed to be effective in the succesive inteference cancellation (SIC) based detector [33].

An example of SAMA matrix $\boldsymbol{S} \in \mathbb{R}^{2\times 3}$ is provided in Fig. 2.2, due to the specific constraint on the ratio $K/N$. With such peculiar design, SAMA was demonstrated to achieve better performance in terms of bit error rate, compared to the orthogonal counterpart [33].

### Multi User Shared Access (MUSA)

Overloading and grant-free access are two key features of MUSA [17]. By adopting grant-free access, MUSA allows each user to choose its spreading sequence freely from a large cardinality, that, in fact, omits the need for resource coordination by the base station. This strategy can reduce signaling overhead and transmission delay caused by conventional grant-based transmission. In this way, MUSA lowers power consumption of devices, and thus is proposed for the uplink transmission [17].

MUSA spreading sequences are required to have short length and low cross-correlation to support a large number of grant-free access users and minimize the impact of user collision. It is, in general, difficult to design a large number of binary spreading codes with low correlation. Non-binary and complex-value random spreading codes with $M$-ary values were proposed for MUSA to address this drawback. Two typical examples of available sets, before normalization, are $\{1+i, -1+i, -1-i, 1-i\}(M=2)$ and $\{0, 1, 1+i, i, -1+i, -1, -1-i, -i, 1-i\}$ $(M=3)$. The entries of these sets are taken from complex spreading codes in the respective $M$-ary set, as shown in Fig. 2.4 for the two respective sets $M=2$ (a) and $M=3$ (b). A MUSA matrix example for $M=3$ is shown in Fig. 2.2.

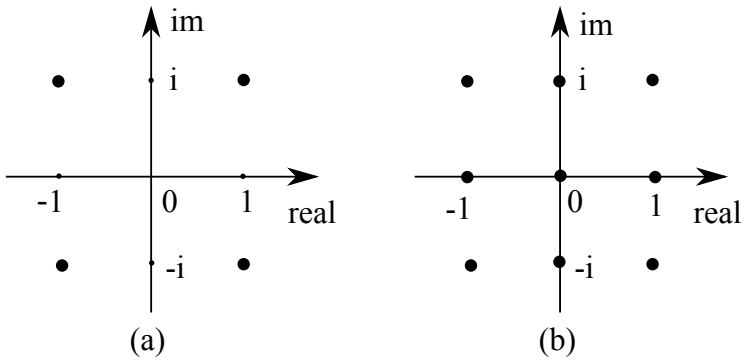



Figure 2.4. Examples of MUSA entries: (a) $M=2$, (b) $M=3$

### Pattern Division Multiple Access (PDMA)

PDMA employs non-orthogonal patterns to map the transmitted data to a resource group, that can be code, power, spatial resource or any combination of those resources, corresponding to code-, power-, or spatial-domain PDMA. In code-domain, PDMA replaces spreading sequences as in LDS-CDMA by non-orthogonal patterns [16], and the number users $K$ are constrained by $2^N - 1$, hence, PDMA is also considered as a special case of SAMA [45]. The concept of PDMA patterns could be understood in a broader way compared to spreading codes. Elements of a PDMA pattern matrix can be either filled with binary numbers in the general

case, or additionally weighted by power scaling and phase shifting in the extended case. Accordingly, the patterns of PDMA assigned to each user admit the following description

$$s_k = \begin{cases} \{0, 1\} & \text{in the general case} \\ \{0, \alpha_{ki} e^{-j\phi_{ki}}\} & \text{in the extended case,} \end{cases}$$

where $\alpha_{ki}$ and $\phi_{ki}$ denote power scaling and phase shifting of the $k^{th}$ user on the $i^{th}$ RE, respectively.

The pattern is designed to enable user separation so that every user has its respective pattern, and as such more users can be multiplexed on a limited number of REs, i.e. is intrisically realized as an overloaded system. In addition, to reduce the receiver complexity, sparsity is also introduced in PDMA patterns. In this way, PDMA is characterized by overloading and sparsity features, and is thus realized as a low-dense system. For example, six users with their different patterns can be mapped to three REs, where one user data can be mapped to one, or even all three REs [16]. An example of the PDMA pattern matrix for $N = 3,\ K = 6$ is

$$\boldsymbol{S}_{\text{PDMA}} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

With the flat fading assumption, the channel matrix is given as $\boldsymbol{H} = \text{diag}\,[h_1, \ldots, h_6]$. The received signal is achieved as

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & 0 & h_4 & 0 & 0 \\ h_1 & 0 & h_3 & 0 & h_5 & 0 \\ 0 & h_2 & h_3 & 0 & 0 & h_6 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}. \tag{2.1}$$

By optimizing power scaling and phase shifting factors in the pattern matrix, PDMA scheme can significantly improve spectral efficiency compared with OMA techniques [16, 28]. The PDMA representation matrix provided in the Fig.2.2 stands for the general case.

### Interleave-grid multiple access (IGMA)

Figure 2.5 shows the schematic of IGMA transmitter, which is characterized by two blocks: bit-level interleavers and grid mapping patterns. Basically, users may be separated by either only bit-level interleaver or grid mapping pattern, or by a combination of both blocks. After channel coding, by applying either low rate FEC code or a combination of a moderate FEC code and repetition code, user data is permuted by a bit-level interleaver [29].

In case only the bit-level interleaver is used to separate users, IGMA becomes IDMA, and the adopted interleaver should be user-specific. After modulation, the generated symbol sequence is subsequently processed by a grid-mapping pattern, including a zero padding block and a symbol-level interleaving. The zero padding

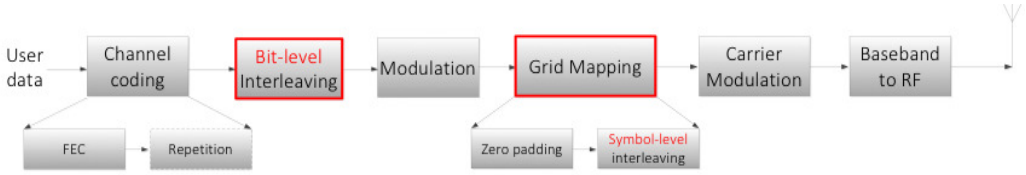



Figure 2.5. Examples of IGMA transmitter [29].

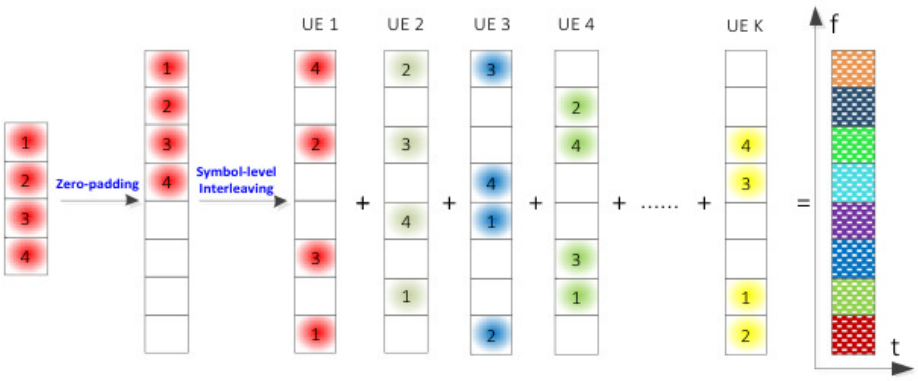



Figure 2.6. Examples of grid mapping process when $N = 4$, $\rho_k = 0.5$ and $L = 8$ [29].

component of the grid-mapping is used in order to reduce the receiver complexity by introducing the sparsity into the mapping of the modulated symbols into the resource elements, while the interleaving is adopted to further randomize interference. An example of grid-mapping process is displayed in Fig. 2.6, where there is simultaneous $K$ users, $N = 8$ bits (chips) and a density $\rho_k = 0.5$, denoting the ratio of used REs (with nonzero element, for example element 1) over the total REs [51].

The spreading matrix $\boldsymbol{S} \in \mathbb{R}^{8\times K}$ of IGMA system model corresponding to the grid mapping example in Fig. 2.6 can be expressed as follows.

$$\boldsymbol{S}_{\text{IGMA}} = \begin{bmatrix} 1 & 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ 1 & 1 & 0 & 1 & \cdots & 1 \\ 0 & 0 & 1 & 0 & \cdots & 1 \\ 0 & 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & 1 & \cdot & 1 \\ \underbrace{1}_{\boldsymbol{s}_1} & \underbrace{0}_{\boldsymbol{s}_2} & \underbrace{1}_{\boldsymbol{s}_3} & \underbrace{0}_{\boldsymbol{s}_4} & \cdots & \underbrace{1}_{\boldsymbol{s}_K} \end{bmatrix}.$$

A summary of aforementioned low-dense SC-NOMA schemes is provided in the Table 2.1, where they are shown to have a mutual link. In particular, SAMA, MUSA, PDMA, and IGMA are shown to be well connected with LDS-CDMA, which all are reflected by sparsity nature of the matrix $\boldsymbol{S}$. It is reasonable then to take LDS-CDMA as *a representative for low-dense NOMA* to compare with the dense case from the

information-theoretic viewpoint.

| **Multiple Access schemes** | *Based-domain* | *Entries of matrix* $\boldsymbol{S}$ | *Relationship with other MA schemes* |
|---|---|---|---|
| **LDS-CDMA/ TH-CDMA** | Sequence-based | $\{+1, 0, -1\}$ | DS-CDMA with low-dense spreading sequences |
| **SAMA** | Sequence-based | $\{0, 1\}$ | LDS-CDMA with specific design of $\boldsymbol{S} \in \mathbb{R}^{N\times(2^N-1)}$ |
| **MUSA** | Sequence-based | $\{1+i, -1+i, -1-i, 1-i\}$ | LDS-CDMA with grant-free access |
| **PDMA** | Pattern-based | $\{0, 1\}$ or $\{0, \alpha e^{-j\phi_{ki}}\}$ | LDS-CDMA with non-orthogonal patterns |
| **IGMA** | Interleave-based | $\{0, 1\}$ | LDS-CDMA with interleavers |

Table 2.1. Properties of low-dense SC-NOMA schemes

## 2.2 Multi-carrier NOMA (MC-NOMA)

### 2.2.1 Dense MC-NOMA

While DS-CDMA is considered as an initial model for single-carrier dense NOMA, multi-carrier CDMA (MC-CDMA) may be inherently referred as a representative for multi-carrier dense NOMA. Basically, MC-CDMA combines DS-CDMA with orthogonal waveforms, for instance OFDM. According to this principle, different dialects for dense multi-carrier NOMA such as multi-carrier resource spread multiple access (MC-RSMA), low code rate and signature based shared access (LSSA), repetition division multiple access (RDMA), group orthogonal coded access (GOCA), are classified into this category and are further described in this section.

#### Multi-carrier Resource Spread Multiple Access (MC-RSMA)

Adopting multi-carrier waveform, the abovementioned RSMA scheme (c.f. Section 2.1) is therefore called as multi-carrier RSMA [43, 44]. It is considered as an optimized solution for low-latency access and it allows for grant-free transmission [44].

#### Low code rate and signature based shared access (LSSA)

LSSA is proposed as a potential candidate for 5G-NR for uplink massive machine-type communication (mMTC) [37]. In this scheme, spreading sequences of DS-CDMA are replaced either by low code rate FEC coding for mitigating multi-user

interference (MUI), or by higher channel coding non-orthogonal spreading sequences [37], and then are multiplexed at bit/symbol level thanks to a particular signature pattern. This signature consists of a reference signal (RS), complex binary sequence, and permutated pattern of a short length vector, which is shared among user's signatures. After multiplexing, the multi-carrier modulation is carried out by IFFT implementation as in OFDM in order to exploit frequency diversity, as shown in Fig. 2.7.

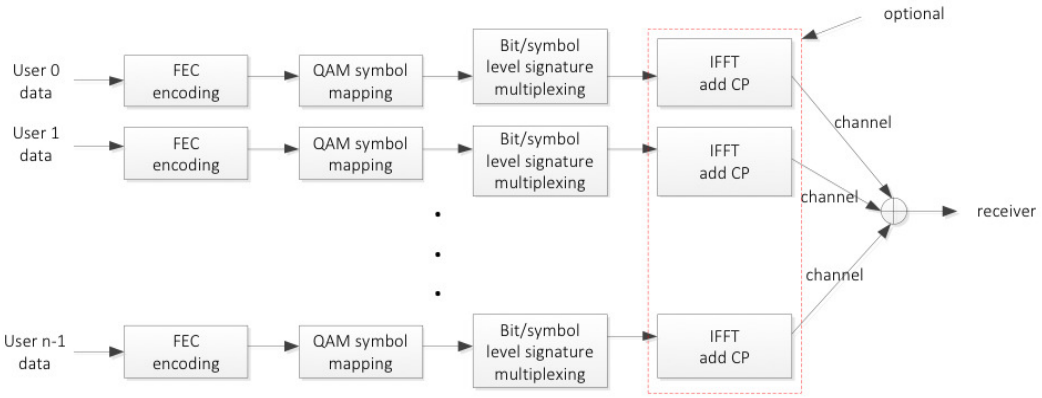



Figure 2.7. Example of LSSA transmitter structure [37]

### Repetition Division Multiple Access (RDMA)

Figure 2.8 shows an example of RDMA transmitter with $K$ users, where a simple cyclic-shift repetition code, instead of spreading code of CDMA, is used to separate users at the receivers and OFDM is employed for multi-carrier waveform [40]. The repetition pattern is designed in such a way to optimize the diversity in both frequency and time for each repeated modulated symbol. In general, RDMA transmitter structure is simpler than MC-RSMA and IDMA since no random interleave or scrambler is required.

### Group Orthogonal Coded Access (GOCA)

While RDMA uses different repetition patterns to distinguish different users, GOCA uses the same repetition pattern for all users, whilst employs an additional two-stage block (see Fig. 2.9). In this block, GOCA sequences are divided into groups, each constitutes a same orthogonal sequence set and different non-orthogonal sequences, that are used subsequently in the first and second stage. Localized time or frequency-repetition, adopted to keep orthogonality among orthogonal sequences in the same group, are expected to significantly reduce multi-user interference. An example of sequence generation block in GOCA transmitter is illustrated in Fig. 2.10.

Two other MC-NOMA candidates proposed by Intel are frequency domain spreading (FDS) and low code rate spreading (LCRS), that function similarly to LSSA and GOCA [42, 52]. Basically, user symbols are spread to multiple resource blocks to

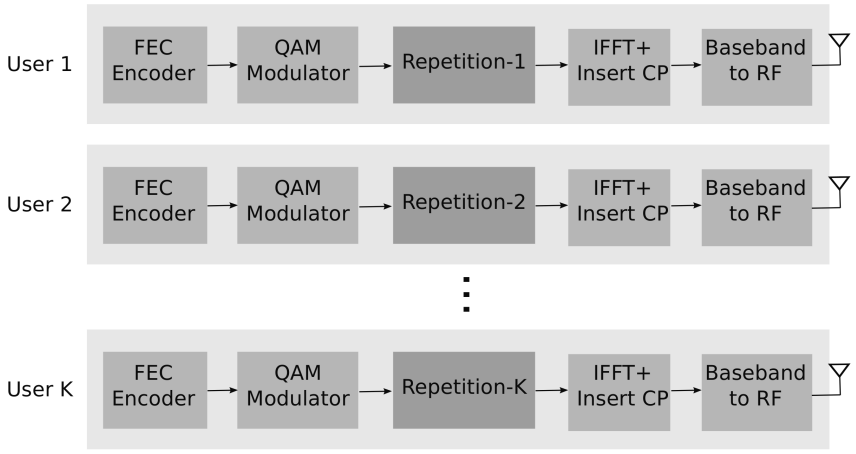



Figure 2.8. Example of RDMA transmitter structure [40]

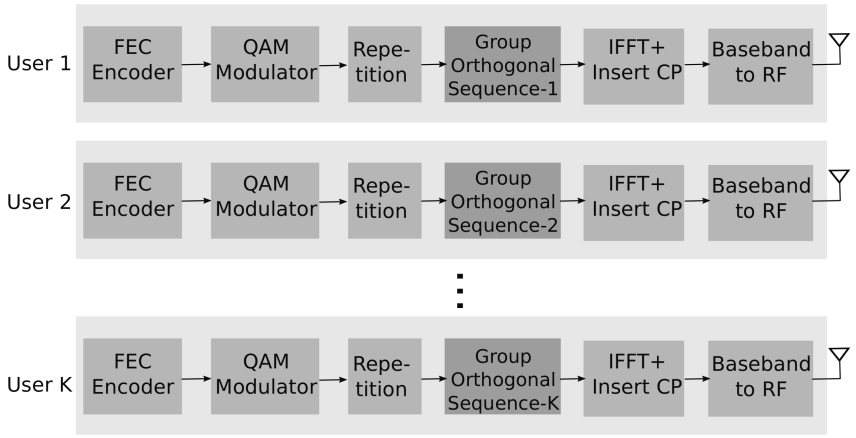



Figure 2.9. Example of GOCA transmitter structure [40]

achieve higher diversity gain, and finally transmitted on multiple waveform, OFDM-based waveform, typically, with the aid of IFFT operation. In addition, user-specific scrambling or interleaving may be employed to further improve multi-user detection performance [42].

### 2.2.2 Low-dense MC-NOMA

In this subclass of MC-NOMA, the low-dense schemes, including LDS-OFDM, SCMA, NCMA, NOCA and LDS-SVE, are described as follows.

#### Low-density spreading OFDM (LDS-OFDM)

LDS-OFDM can be seen as a combined system of LDS-CDMA and OFDM, where LDS-CDMA is applied as multiple access technique and OFDM is applied for multi-carrier modulation. In particular, data symbol of each user is spread over a small set of sub-carriers due to the low-dense property of signatures. Each sub-carrier is shared by many users, hence, is designed to be orthogonal and used by a limited number of

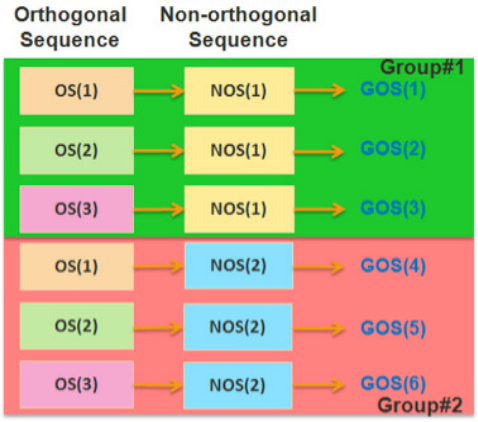



Figure 2.10. Example of sequence generation in GOCA transmitter structure [40]

symbols to avoid superposition [35]. By adopting spreading sequences, LDS-OFDM is, in fact, an enhanced version of MC-CDMA, where the difference is that the spreading signatures are sparse or low-dense by employing zero padding [35].

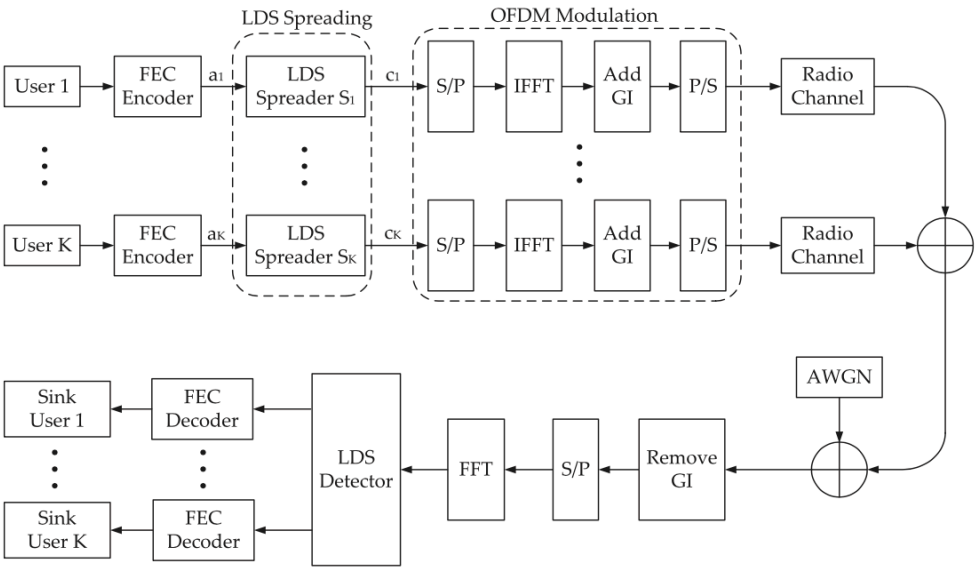



Figure 2.11. Example of Uplink MC-LDSMA (LDS-OFDM) block diagram [53]

Very first concept of LDS-OFDM came from Jinho Choi [34] termed as low-density spreading for multicarrier system (LDS-MC). In [34], the number of nonzero elements $N_s$ in the spreading sequences was considered as an evaluator for the tradeoff between diversity gain and computational complexity of receivers. The diversity gain grows proportional to $N_s$, whereas the complexity also increases. Maximum Likelihood detection and iterative receiver (including a MAP detector and decoder) were employed for uncoded and coded LDS-MC systems, respectively.

For both cases, the performance improvement in terms of BER has been observed compared to conventional OFDM. Later on, LDS-OFDM, termed as Multicarrier low-density spreading multiple access (MC-LDSMA) in [53], [35], was studied with different proposals at the receiver side. An example of MC-LDSMA block diagram is shown in Fig. 2.11. A turbo multi-user detector/decoder (MUDD) using extrinsic information transfer (EXIT) charts was proposed in [14] was shown to gain the BER improvement with less iterations, compared to the same scenario of MC-CDMA.

Based on bipartite graph, a joint sparse graph for OFDM, including single graph of LDS-OFDM and LDPC codes, was proposed in [35]. At the receiver, a corresponding joint multiuser detection with FEC decoding was presented, showing an enhancement in terms of BER over conventional systems such as group-orthogonal multi-carrier CDMA, LDS-OFDM and turbo structured LDS-OFDM. Due to the constraints on the number of users sharing the same subcarrier and the number of subcarrier for spreading each symbol, radio resource allocation for LDS-OFDM naturally becomes a challenge for 5G system design. In [53], proposed power-allocation algorithms have been considered in single-user and multiuser scenarios. Optimal and suboptimal algorithms have been proposed aiming to solve the two following tasks: sub-carrier partition to maximize the user rate (water-filling) and optimizing power allocation for a given sub-carrier partitioning (a maximum ratio transmission) [53]. Numerical analysis showed that proposed algorithms improved performance in terms of spectral efficiency and outage probability compared to that of OFDMA.

Related to bipartite graph, it is also worth to mention random access schemes, in parallel with CDM-NOMA, built on the combination of packet erasure correcting codes and SIC technique with ALOHA [54, 55]. This may enable the possibility to use coding theory for designing efficient random access protocols, which are of practical interests, particularly in IoT scenarios.

### Sparse Code Multiple Access (SCMA)

SCMA is an extended case of LDS-CDMA, where modulation and spreading are combined and merged into a unified process, named SCMA encoder [36]. In particular, the SCMA encoder, first maps the encoded bits into $K$ complex vectors (codewords), corresponding to $K$ users, then these vectors are multiplexed over $N$ orthogonal resources, for example OFDMA, and transmitted on the radio channels. Since SCMA maps those encoded bits directly to sparse multi-dimensional codewords, that are equivalent to joint processing of modulation and spreading, this brings further a 'shaping gain' from optimization of a multi-dimensional constellation. An example of SCMA encoding and multiplexing is shown in Fig. 2.12.

To distinguish different users and avoid collision among users, codewords, dedicating for a user, are selected from a predefined codebook set, and are, typically, sparse, complex, and multi-dimensional. This feature makes the decoding process time consuming, as dictated as an exponential order in computational complexity. Since factor graph is exploited for presenting the mapping matrix, entries of mapping matrix $\boldsymbol{S}$ consist of binary values, indicating the index of the nonzero elements of signal input.

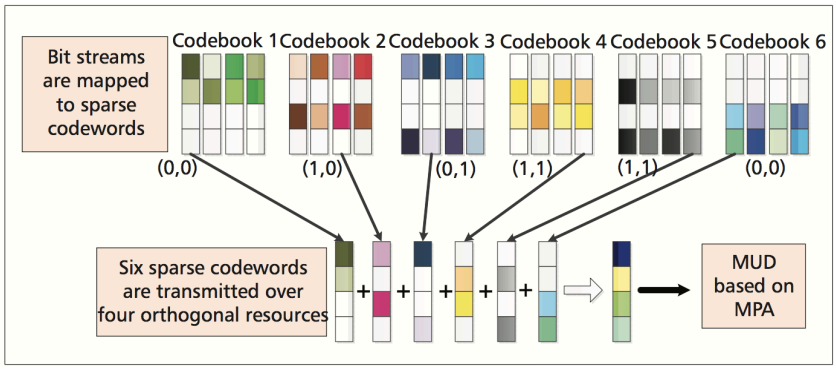



Figure 2.12. Example of SCMA encoding and multiplexing [33]

### Non-orthogonal coded multiple access (NCMA)

Similar to SCMA, NCMA also adopts the non-orthogonal codebooks, whose codewords obtained by Grassmannian line packing problem [38]. As illustrated in Fig. 2.13, receivers distinguish different users with the aid of UE specific non-orthogonal code cover (NCC), which represent a non-orthogonal codeword attached to each user.

By employing additional layers via superposed symbol, the throughput of NCMA increases at an expense of small BLER loss under specific scenarios, for example, huge connections with small packet in mMTC, or for reducing the collision probability in contention based MA. As a matter of fact, NCMA is favorable due to low receiver complexity by adopting parallel interference cancellation (PIC) at the receiver.

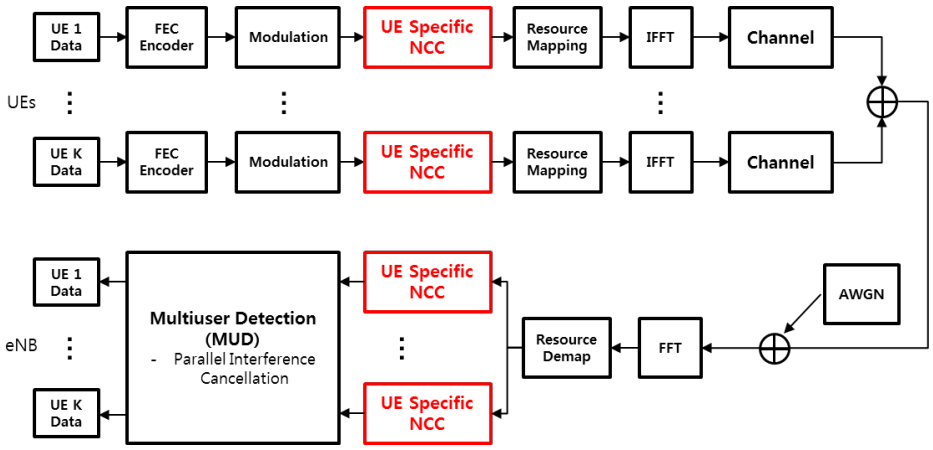



Figure 2.13. Example of NCMA structure in UL [38]

### Non-orthogonal coded access (NOCA)

Figure 2.14 shows an example of NOCA transmitter, where SF is the spreading factor and $C_{\mathrm{SF}}^{j}$ is the spreading sequence of the $j$-th user. As other proposed NOMA

schemes, the NOCA idea is that data symbols are spread in either frequency or time domain by non-orthogonal sequences before transmission [39].

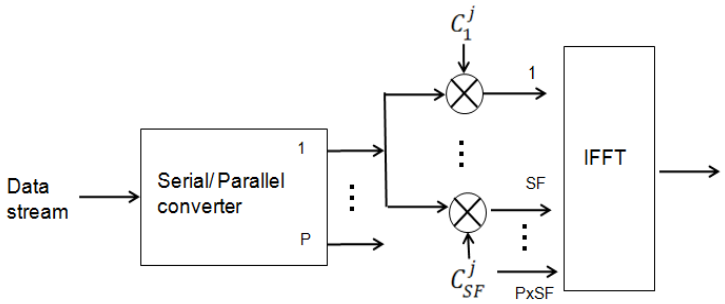



Figure 2.14. Example of NOCA transmitter structure [39]

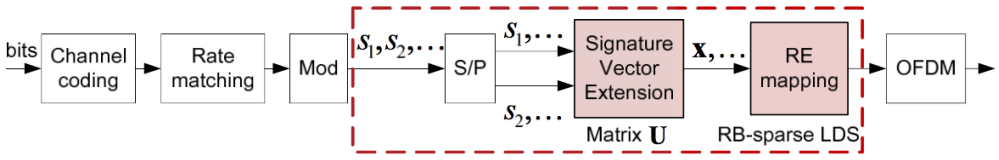



Figure 2.15. Example of LDS-SVE transmitter structure [41]

The original modulated data stream is first converted into $P$ parallel sequences, each is then spread to SF subcarriers, leading to $P \times$ SF subcarriers in total. NOCA sequences are designed to have constant modulus, low cubic metric, and low correlation properties as dictated in [39].

### Low Density Spreading-Signature Vector Extension (LDS-SVE)

LDS-SVE is an extended version of LDS-CDMA using multi-carrier waveform (two-carrier in particular), proposed by Fujitsu for uplink OFDM system [41]. The transmitter structure of LDS-SVE illustrated in Fig. 2.15 shows that after modulation, each symbol is spread onto the two subcarriers. This is obtained with the aid of transformation matrix $\boldsymbol{U}$ in multiplication with the real and imaginary parts of two modulate symbols, denoted by $s_1$, $s_2$. The output forms a vector $\boldsymbol{x}$ constituting two parallel symbols, and is further mapped onto REs sparsely, as in LDS-CDMA case. The mapped subcarriers are, in general, designed in an optimized way to increase the diversity gain.

# Chapter 3

# Theoretical analysis of SC-NOMA over AWGN channels

## 3.1 Theoretical framework

In this chapter, the theoretical behavior of SC-NOMA is analyzed under the impact of three main factors. First, the load factor provides a straightforward way to study the system behavior in the underloaded ($\beta < 1$) vs. overloaded ($\beta > 1$) regimes. As mentioned above, overloaded systems are necessarily NOMA, since as soon as $\beta$ overcomes the boundary value $\beta = 1$, new users find all REs occupied. In fact, SC-NOMA [1] schemes introduced in Chapter 2 are characterized by overloaded property, i.e. $\beta > 1$.

The second typical feature under analysis is the dense vs. low-dense aspect, corresponding to the dense vs. low-dense groups. Either all $N$ dimensions are used, as in CDMA or IDMA, or only a part of the dimensions is used, as in low-dense systems. Due to sparsity, the system can be graded from dense to extreme low-dense. The number of used dimensions, that we call $N_s$, defines the *degree of sparseness*. If $N_s = N$, the system is dense. If $N_s = 1$, then the system is extreme low-dense. All other degrees of sparseness lie in between these two extreme cases. The dense vs. low-dense feature is directly reflected by the properties of matrix $\boldsymbol{S}$, as defined in the mathematical model (eq. (1.1)). Regarding the energy feature, the matrix $\boldsymbol{S}$ of low-dense system contains most of elements that are '0', where '0' indicates elements with zero energy. A heuristic way to think of NOMA scheme is thus as a version of the overloaded CDMA scheme and low-dense NOMA can be referred to as sparse overloaded CDMA [47]. Naturally, it is expected to investigate the effect of those NOMA parameters, including the load $\beta$ and the degree of sparseness $N_s$, on theoretical behavior of dense vs. low-dense NOMA.

Achievable rates of low-dense NOMA in the LSL were early and extensively evaluated via sparse CDMA by means of the replica method, also known as heuristic statistical physics, in [47–50]. Since the derivations provided by replica method were typically non-rigorous, the information-theoretic analyses on low-dense NOMA were found rigorously via closed-form expressions in LDS/TH-CDMA model [21, 32],

[1] afterwards SC-NOMA is called shortly as NOMA in this chapter

and in regular sparse NOMA model [22, 56]. Given that multiple system models are possibly proposed due to different assumptions, below we reported all curves along with the existing relevant theoretical results in our analytical model (c.f. Chapter 1).

Regarding the third feature, the regularity, low-dense NOMA ($N_s < N$) are further classified into *irregular* vs. *regular* based on spreading mapping constraints, given that $N_s$ is also the number of *occupied* REs per user, whereas $N$ is the total number of REs per user. Previous works on sparse CDMA and low-dense NOMA were classified as *irregular* since the number of occupied REs per user was randomly Poissonian distributed with fixed mean [47, 48], and randomly uniformly distributed [21, 32][2], respectively. On the other hand, in terms of spreading matrix, the regularity assumption in [22, 56] requires matrix $\boldsymbol{S}$ be structured with exactly $N_s \in \mathbb{N}^+$ and $\beta N_s \in \mathbb{N}^+$ non-zero entries per column and row, respectively. It is equivalent to having each user occupying $N_s$ REs and each RE being allocated with exact $\beta N_s$ users, subject to $N_s$ and $\beta N_s$ being integers. It is, in general, challenging to have such an ideal model in practical scenarios where users are not allowed to independently select the spreading sequences, they must be coordinated or central scheduled [56]. The *regular* low-dense NOMA via *regular* sparse CDMA was early demonstrated to be superior to the dense in terms of bit error rate in high noise regime in [49], and in terms of spectral efficiency via explicit analytical expressions in recent works [22, 56].

In the following, theoretical behavior of *irregular* vs. *regular* low-dense NOMA ($N_s < N$) is analyzed with the adopted reference models LDS/TH-CDMA [21, 32] vs. regular sparse NOMA [22, 56], respectively. DS-CDMA is adopted as a representative of the dense NOMA group ($N_s = N$) [20]. Both optimal and linear receivers are considered in all cases. Spectral efficiency expressions [bits/s/Hz] for different cases are reported for the self-contained purpose of the chapter.

It is important to notice that the theoretical results of *irregular* low-dense NOMA are available only for $N_s = 1$ [21, 32] (see 3.2), while closed-form expressions of the *regular* case are valid only for intermediate degrees of sparseness, specifically for $N_s \geqslant 2, \beta N_s \in \mathbb{N}^+$ [22] (see 3.3). For *irregular* low-dense NOMA, since the closed-form expressions for intermediate $N_s$ do not exist yet in the literature (and in general, are not easy to achieve), the results will be shown via Monte Carlo simulations for a full coherent overview. For *regular* low-dense NOMA, the regularity in case of $N_s = 1$ yields a typical setting, which includes a set of parallel Gaussian multiple access channels (MAC), that will be investigated further in 3.3. The references for mapping information-theoretic results in code-domain NOMA are summarized in Table 3.1, with the corresponding numbered equations in this thesis.

## 3.2 Dense vs. *Irregular* low-dense NOMA

In this part, theoretical behavior of dense vs. *irregular* low-dense NOMA is analyzed with the two corresponding reference models, that are DS-CDMA ($N_s = N$) [20] and LDS/TH-CDMA ($N_s < N$) [21]. Since the AWGN channel is used for both cases,

[2]The irregular low-dense NOMA in [21, 32] is called as *partly-regular* sparse NOMA in [22, 56]

| | Dense NOMA ($N_s = N$) | Low-dense NOMA ($1 < N_s < N$) | | Low-dense NOMA ($N_s = 1$) | |
|---|---|---|---|---|---|
| | | Irregular | Regular | Irregular | Regular |
| Optimum receivers | Eq. (3.2) [20] | Eq. (3.1) by Monte-Carlo simulations [20] | Eq. (3.9) [22] | Eq. (3.4) [21] | Eq. (3.11) of this thesis |
| Linear receivers | Eqs. (3.5, 3.6) [20] | Eq. (3.7) [21] | Eq. (3.12) [22] | Eq. (3.8) [21] | Eq. (3.13) of this thesis |

Table 3.1. Summary of available theoretical bounds with corresponding references

the channel matrix $\boldsymbol{H}$ in eq. (1.1) becomes an identity matrix. The *only difference* in the mathematical model between DS-CDMA and LDS-CDMA is situated in the sparseness of matrix $\boldsymbol{S}$. In DS-CDMA ($N_s = N$), all entries of $\boldsymbol{S}$ are randomly filled by binary values of $\{\pm 1\}$, while in LDS-CDMA, for example with $N_s = 1$, each column of $\boldsymbol{S}$, representing a user, contains only *one* nonzero entry ($\{+1\}$ or $\{-1\}$), and all the rest are nil.

### 3.2.1 Optimum receivers

The general spectral efficiency for optimum receivers in both dense NOMA and low-dense NOMA cases can be computed via [20–22]

$$C_{\text{opt}}^{N}(\gamma) = \frac{1}{N} \log_2 \det [\boldsymbol{I} + \gamma \boldsymbol{S}\boldsymbol{S}^*], \tag{3.1}$$

where $\gamma$ denotes the signal-to-noise (SNR) ratio.

#### Dense NOMA

The optimum capacity for the dense case is equal to [20]

$$\begin{aligned} C_{\text{opt}}^{\text{dense}}(\beta, \gamma) = {} & \frac{\beta}{2} \log_2 \left(1 + \gamma - \frac{1}{4}\mathcal{F}(\gamma, \beta)\right) \\ & + \frac{1}{2} \log_2 \left(1 + \gamma\beta - \frac{1}{4}\mathcal{F}(\gamma, \beta)\right) \\ & - \frac{\log_2 e}{8\gamma}\mathcal{F}(\gamma, \beta), \end{aligned} \tag{3.2}$$

where

$$\mathcal{F}(x, z) \coloneqq \left(\sqrt{x(1 + \sqrt{z})^2 + 1} - \sqrt{x(1 - \sqrt{z})^2 + 1}\right). \tag{3.3}$$

#### *Irregular* low-dense NOMA

- $1 < N_s < N$: In order to study the behavior of the system, Monte-Carlo simulations are introduced for optimum receivers of *irregular* low-dense NOMA based on eq. (3.1).

- $N_s = 1$: The closed-form capacity expression of the *irregular* case is [21]

$$C_{\text{opt}}^{\text{irreg}}(\beta,\gamma) = \sum_{k\geqslant 0} \frac{\beta^k e^{-\beta}}{k!} \log_2(1+k\gamma)\,. \tag{3.4}$$

It is worth mentioning that, Maximum Likelihood is typically adopted as the optimum receiver and is, in general, favorably replaced in real-world communication systems by sub-optimal receivers due to complexity issue.

### 3.2.2 Linear receivers

For dense NOMA, closed-form capacity expressions for all linear receivers are available as in [20]. For *irregular* low-dense NOMA, while the same mutual information is obtained for different linear receivers, including minimum mean square error (MMSE), single user matched filter (SUMF), and zero forcing (ZF) with respect to the particular case $N_s = 1$, closed-form achievable rates for $1 < N_s < N$ are available only for the SUMF receiver [21]).

#### Dense NOMA

The spectral efficiency of SUMF and MMSE receivers for the dense system in the LSL is [20]

$$C_{\text{SUMF}}^{\text{dense}}(\beta,\gamma) = \frac{\beta}{2}\log_2\left(1+\frac{\gamma}{1+\gamma\beta}\right), \tag{3.5}$$

and

$$C_{\text{MMSE}}^{\text{dense}}(\beta,\gamma) = \frac{\beta}{2}\log_2\left(1+\gamma-\frac{1}{4}\mathcal{F}(\gamma,\beta)\right), \tag{3.6}$$

where $\mathcal{F}(\gamma,\beta)$ is defined as in eq. (3.3).

#### *Irregular* low-dense NOMA

Due to the peculiar sparse structure of matrix $\boldsymbol{S}$ in LDS-CDMA when $N_s = 1$, the linear receivers (MMSE, SUMF, ZF) yield the same mutual information, while it remains unknown for the in-between cases $1 < N_s < N$ for the MMSE and ZF receivers [21]. Therefore, theoretical behavior of *irregular* low-dense NOMA linear receivers for the intermediate degree of sparseness cases is investigated via the mutual information of the SUMF receiver of LDS-CDMA [21] as follows:

- $1 < N_s < N$: The mutual information of *irregular* low-dense NOMA with the SUMF receiver writes [21]

$$R_{\text{lin}}^{\text{irreg}}(\beta,\gamma,N_s) = \beta\sum_{k\geqslant 0}\frac{(N_s^2\beta)^k}{k!}e^{-N_s^2\beta}\log_2\left(1+\frac{\gamma}{1+\frac{k}{N_s}\gamma}\right). \tag{3.7}$$

- $N_s = 1$: The closed-form expression of the *irregular* low-dense achievable rate with linear receivers when $N_s = 1$ is [21]

$$R_{\text{lin}}^{\text{irreg}}(\beta,\gamma) = \beta\sum_{k\geqslant 0}\frac{\beta^k e^{-\beta}}{k!}\log_2\left(1+\frac{\gamma}{k\gamma+1}\right). \tag{3.8}$$

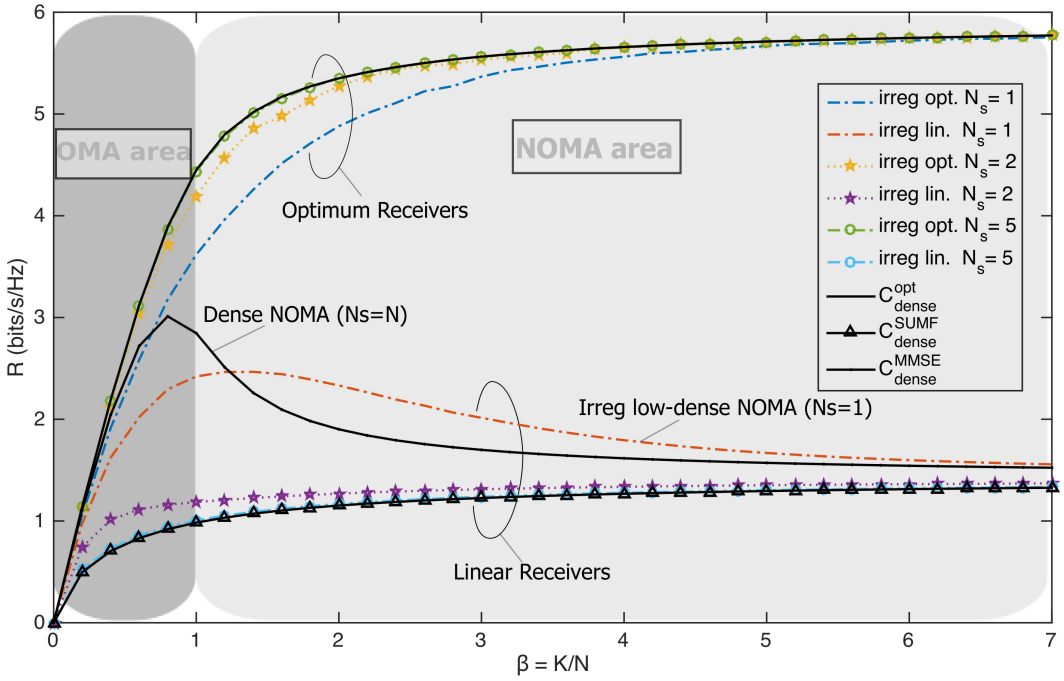



Figure 3.1. Achievable rates (bits/s/Hz) of dense NOMA vs. *irregular* low-dense NOMA as a function of $\beta$ with fixed $E_b/N_0 = 10$ [dB] (parts of the data used to draw this figure were extracted from [21])

Figure 3.1 shows the achievable rates of dense vs. *irregular* low-dense systems with optimum and linear receivers as a function of $\beta$ with fixed value of $E_b/N_0 = 10$ [dB]. With respect to load factor $\beta$, the border line (vertical dashed line) at $\beta = 1$ divides Fig. 3.1 into two areas corresponding to OMA (underloaded with $\beta < 1$) and NOMA (overloaded with $\beta > 1$), with dark and light shaded area, respectively.

In the LSL, Fig. 3.1 shows that for optimum receivers, dense systems always outperform *irregular* low-dense, irrespective of $\beta$, that is, whether OMA or NOMA. Achievable rates for the *irregular* type drop with $N_s$ from the dense case ($N_s = N$) to the extreme low-dense case ($N_s = 1$), and the gap between the *irregular* low-dense and dense becomes negligible at $N_s = 2$, and tends to vanish from $N_s > 2$, e.g. $N_s = 5$. On the other hand, the behavior of linear detection changes, with respect to the level of density of the system. For MMSE receivers, achievable rates of the dense systems are higher than the *irregular* low-dense in the OMA area, while this situation is inverse in the NOMA area, starting from about $\beta > 1.2$. With growing $N_s$, for example $N_s = \{2, 5\}$, the gap between the achievable rates of *irregular* low-dense NOMA with the SUMF receiver and dense NOMA, sharply reduces, to converge to the SUMF dense curve. Given that optimum detection is unfeasible to implement in practice due to the receiver complexity, the above observation provides the ground for suggesting *irregular* low-dense NOMA in the LSL, for example, for *irregular* low-dense case with $N_s = 1$.

The reported analysis holds in the case of flat-fading channel, as investigated and proved in [32].

$$\mathcal{G}(x,y,z) \coloneqq \left( \frac{\sqrt{(y-(1-\sqrt{z})^2)(x(1+\sqrt{z})^2+1)} - \sqrt{(y-(1+\sqrt{z})^2)(x(1-\sqrt{z})^2+1)}}{\sqrt{y-(1-\sqrt{z})^2} - \sqrt{y-(1+\sqrt{z})^2}} \right)^2,$$
$$x,y,z \in \mathbb{R}^+, y \geqslant (1+\sqrt{z})^2. \tag{3.10}$$

## 3.3 Dense vs. *Regular* low-dense NOMA

Theoretical analysis of *regular* low-dense NOMA is investigated in this part, in comparison with the dense case.

- $1 < N_s < N$: Closed-form expressions of *regular* low-dense NOMA achievable rates for both optimum and MMSE receivers in [22] are valid under the following constraints:
  - if each user has $2 \leq N_s \in \mathbb{N}^+$ non-zero entries in its spreading sequence, $2 \leq \beta N_s \in \mathbb{N}^+$ users shoud be assigned in the same RE;
  - spreading matrix $\boldsymbol{S}$ is assumed to converge to a bipartite Galton-Watson tree in the LSL (see ( [22], Theorem 2) for a full description).
  To effectively induce non integer values of $N_s$ and $\beta N_s$, one may employ time-sharing between different $(N_s, \beta N_s)$ points in the admissible set to achieve the same total throughput as mentioned in ( [22], Remark 4).
- $N_s = 1$: The regularity imposes $\beta \in \mathbb{N}^+$ users per each RE, that is equivalent to having a set of $N$ parallel Gaussian MAC channels [57]. This observation may bring more insight on the behavior of the *regular* with respect to optimum and linear receivers.

### 3.3.1 Optimum receivers

- $1 < N_s < N$: For the *regular* one, the closed-form expression of optimum capacity is valid for $N_s \geqslant 2$, subject to $N_s, \beta N_s$ being integers [22]

$$\begin{aligned} C_{\text{opt}}^{\text{reg}}(\gamma,\beta,N_s) = {} & \frac{\beta(N_s-1)+1}{2} \log_2\left(1+(\delta+\alpha)\gamma - \frac{1}{4}\mathcal{F}(\delta\gamma,\tilde{\beta})\right) \\ & + (\beta-1)\log_2\left(1+\alpha\gamma-\frac{1}{4}\mathcal{F}(\delta\gamma,\tilde{\beta})\right) \\ & - \frac{\beta(N_s-1)-1}{2}\log_2\left(\frac{(1+\beta N_s\gamma)^2}{\mathcal{G}(\delta\gamma,\psi,\tilde{\beta})}\right), \end{aligned} \tag{3.9}$$

  where $\alpha \coloneqq 1 - 1/N_s,\ \delta \coloneqq \beta - 1/N_s,\ \tilde{\beta} \coloneqq \alpha/\delta,\ \psi \coloneqq \beta N_s/\delta$,
  $\mathcal{F}(x,z)$ defined in (3.3) and $\mathcal{G}(x,y,z)$ defined as in eq. (3.10).
- $N_s = 1$: In this case, the *regular* low-dense NOMA scheme turns out a set of $N$ parallel Gaussian MAC channels, each with $\beta$ ($1 < \beta \in \mathbb{N}$) users. This

setting makes sense only in the overloaded regime. One may easily obtain the spectral efficiency that in this case is

$$C_{\text{opt}}^{\text{MAC}}(\beta,\gamma) = \log_2(1+\beta\gamma). \tag{3.11}$$

It is worth highlighting that finding the spectral efficiency as a function of $E_b/N_0$ (via the relation $E_b/N_0 = \beta\gamma/C(\gamma)$ [19]) yields exactly the Cover-Wyner bound as in [20], corresponding to a system with no spreading.

The behavior of *regular* low-dense NOMA vs. dense NOMA with optimum receivers is shown on Fig. 3.2. For the sake of full comparison, theoretical behavior of the *irregular* low-dense with $N_s = 1$ and the orthogonal case are also plotted for reference. Achievable rates are plotted as a function of system load $\beta$, for a fixed value of $E_b/N_0 = 10\,[\text{dB}]$ (Fig. 3.2a), and as a function of $E_b/N_0$, for fixed $\beta = 2$ (Fig. 3.2b). In contrast to the *irregular* counterpart, achievable rates for the *regular* low-dense NOMA, which are superior to all other cases, grow gradually for lower values of $N_s < N$, and reach the ultimate rate (Cover-Wyner bound) when $N_s = 1$.

The reason that makes the optimal spectral efficiency of the *irregular* low-dense to be lower than the dense case may be caused by the random nature of user-resource allocation, leading to a condition in which some users are not assigned with any RE, while some REs are left unused. On the other hand, the regularity feature of the *regular* low-dense NOMA contributes to increasing the optimal spectral efficiency by employing user-mapping intentionally. Nonetheless, this also imposes as a direct consequence additional practical challenges in having some kind of coordination while allocating the resources to users [22, 56].

### 3.3.2 Linear receivers

- $1 < N_s < N$: The closed-form expressions of *regular* low-dense NOMA achievable rates for linear MMSE (LMMSE) are [22]

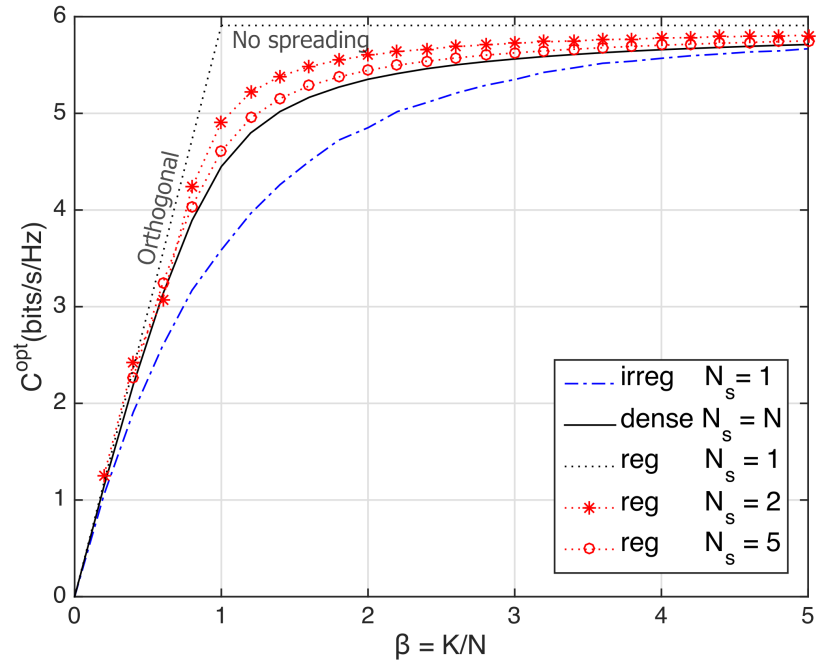
$$C_{\text{LMMSE}}^{\text{reg}}(\gamma,\beta,N_s) = \beta\log_2\left(\frac{1+\beta N_s\gamma}{1+N_s\delta\gamma-\frac{N_s\mathcal{F}(\delta\gamma,\tilde{\beta})}{4}}\right), \tag{3.12}$$

where $\delta$, $\tilde{\beta}$, $\mathcal{F}(x,z)$ defined as in eq. (3.9).

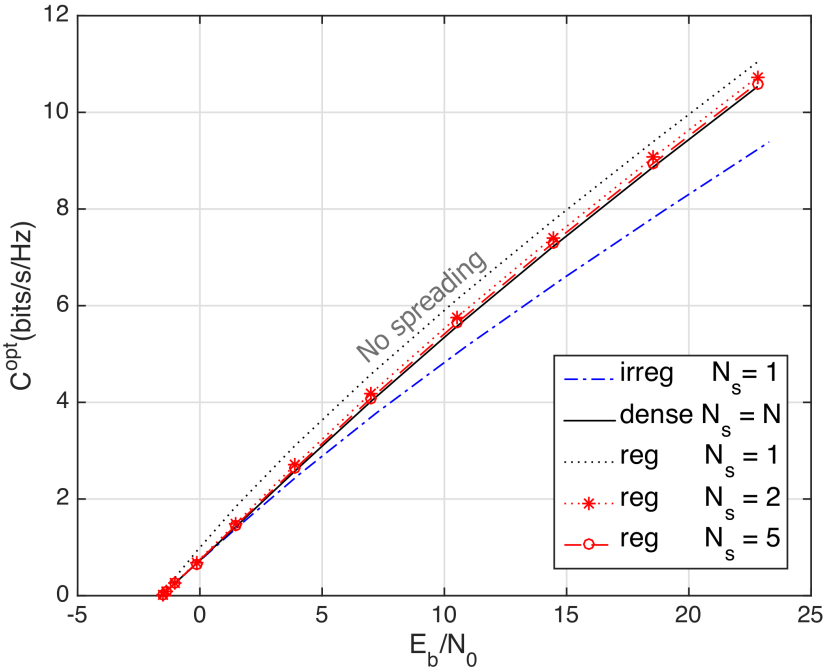
- $N_s = 1$: In this particular case, the achievable rate of the *regular* low-dense NOMA with respect to the LMMSE receiver may be derived via the setting of parallel MAC channels, each with $\beta$ users as follows

$$R_{\text{LMMSE}}^{\text{MAC}}(\gamma,\beta) = \beta\log_2\left(1+\frac{\gamma}{1+(\beta-1)\gamma}\right). \tag{3.13}$$

Figure 3.3 shows the achievable rates of MMSE, ZF, SUMF receivers for the dense ($N_s = N$), LMMSE receiver for the *regular* low-dense schemes with $N_s = \{1, 2, 5\}$. The linear receiver for the *irregular* low-dense NOMA with the typical case $N_s = 1$ is also shown for comparison. A remarkable observation from low-dense NOMA with $N_s = 1$ can be given: capacity of *regular* low-dense NOMA with linear receiver outperforms all the rest when $\beta \leqslant 1$ (OMA area), particularly to the typical setting




(a)




(b)

Figure 3.2. Achievable rates (bits/s/Hz) of dense NOMA ($N_s = N$) vs. *regular* and *irregular* low-dense NOMA ($N_s < N$) with optimum receivers
(a) as a function of $\beta = K/N$ for fixed $E_b/N_0 = 10$ [dB]
(b) as a function of $E_b/N_0$ for fixed $\beta = 2$

when $N_s = 1$; while in the overloaded regime (NOMA area), there is an intersection where capacity of *irregular* low-dense NOMA with $N_s = 1$ outperforms all other cases. By numerical equation solving, the exact value of the intersection is located at $\beta = 1.232$, from which *irregular* low-dense NOMA with $N_s = 1$ dominates those of dense NOMA (c.f. Figs. 3.1 and 3.2), as well as with all other degrees of sparseness ($N_s > 1$) till about $\beta \approx 5$, and then tend to converge for $\beta > 5$ (with the negligible gap of about 5% at $\beta = 5$). These observed results can be used as a driving rationale in system design.

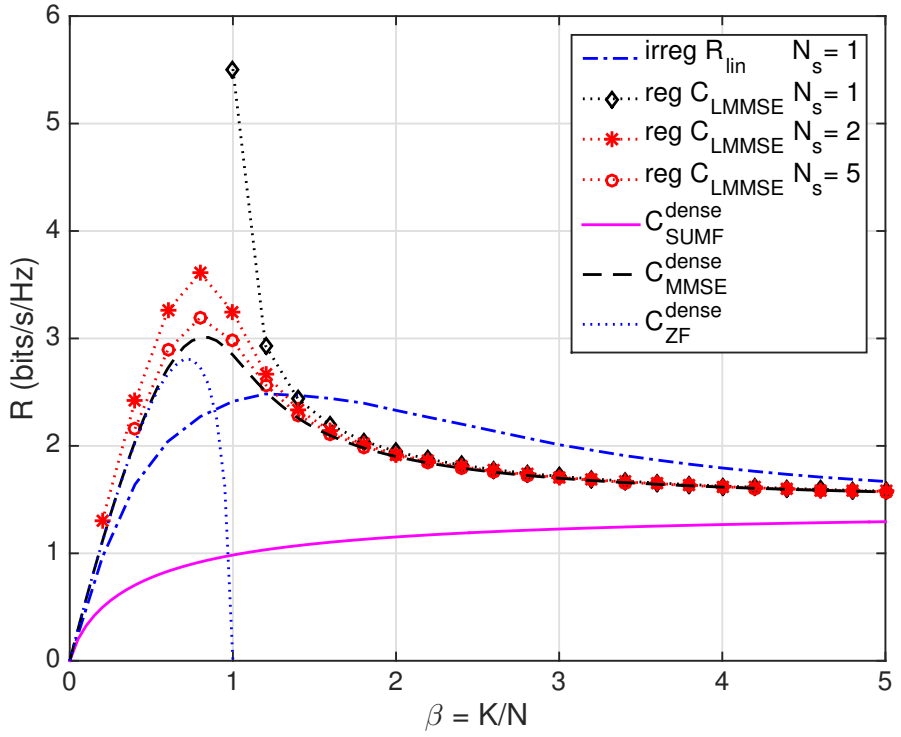



Figure 3.3. Achievable rates (bits/s/Hz) of dense NOMA ($N_s = N$) with SUMF, MMSE, ZF receivers vs. low-dense NOMA ($N_s = 1, 2, 5$) with linear receivers as a function of $\beta$ for fixed $E_b/N_0 = 10$ [dB]

## 3.4 Conclusion

Motivated by the key challenge of finding and analyzing theoretical bounds for NOMA in massive communications, this chapter sheds some light on the relationship between achievable rates and NOMA parameters, such as load factor, degree of sparseness and regularity.

Theoretical investigations were interpreted in the LSL for both optimum and linear receivers, based on closed-form expressions existing for three distinctive cases, that are, dense vs. *regular* low-dense and *irregular* low-dense NOMA, corresponding to DS-CDMA ($N_s = N$) [20] vs. LDS-CDMA ($N_s = 1$) [21] and *regular* sparse

NOMA ($2 \leq N_s \in \mathbb{N}^+, \beta N_s \in \mathbb{N}^+$) [22]. For any value of load, low-dense NOMA cases were shown to be more spectral-efficient than dense ones. For optimum receivers, achievable rates of the *regular* low-dense are higher than the *irregular* low-dense and dense NOMA regardless of load. To this end, the system must be constrained to have exactly $N_s$ REs per user and $\beta N_s \in \mathbb{N}$ users per resource; this imposes either central scheduling or users coordination. For linear receivers, spectral efficiency of *regular* low-dense NOMA was proved to be higher than all the other cases in the underloaded regime, while spectral efficiency of *irregular* low-dense dominated other NOMA cases in the overloaded systems, particularly when the system load $\beta$ is within an interval that is about $[1.2, 5]$. When $N_s$ increases, that is sparseness decreases, achievable rates of low-dense cases rapidly converged to achievable rates of the dense case, as soon as $N_s = 2$.

In conclusion, by changing the spreading strategy from dense to low-dense, specific theoretical limits hold, showing that, to obtain higher achievable rates for linear decoders while still enjoying the lower receiver complexity, it is advisable to adopt sparse communications, and in particular *irregular* extreme low-dense schemes when systems are overloaded and *regular* extreme low-dense cases in the underloaded regime.

# Chapter 4

# Theoretical analysis of Low-dense SC-NOMA over Rayleigh fading channels with perfect CSI

In this chapter, spectral efficiency of low-dense SC-NOMA in the presence of fading with perfect knowledge of channel is derived for linear detection with independent decoding as well as optimum decoding. In the case of optimum decoding, it is found that low-density spreading underperforms dense spreading for all loads. Conversely, linear detection is characterized by different behaviors in the underloaded vs. overloaded regimes. In particular, it is shown that spectral efficiency changes smoothly as load increases. However, in the overloaded regime, the spectral efficiency of low-density spreading is higher than that of dense spreading.

## 4.1 Introduction

### 4.1.1 Background and Motivation

In SC-NOMA, it is characterized by different dialects, such as low-density spreading CDMA (LDS) [11–13], multi-user shared access (MUSA) [17], successive iterative cancellation amenable multiple access (SAMA) [33], pattern division multiple access (PDMA) [16, 28], interleave division multiple access (IDMA) [30, 31] and interleave-grid multiple access (IGMA) [29] and so on. As a matter of fact, those NOMA variants enable flexible resource allocation, and reduce hardware complexity by relaxing orthogonality requirements.

In this work, we focus on LDS. As a typical variant of SC-NOMA, LDS inherits all above advantages and will be shown later in this chapter to obtain increased system throughput compare to conventional CDMA, particularly in massive communications. LDS, therefore, may be appropriately fit to IoT scenario [2] and is also considered as a potential candidate for uplink machine-type-communications (mMTC) [2]. Conventional direct-sequence CDMA (DS) is based on the spread spectrum technique, that uses spreading sequences to spread the signal over a given

bandwidth. In traditional CDMA, signal dimensions, also known as chips (the terminology stemmed from the chip-rate of the sample), are all filled in with nonzero values, making the structure of DS be a form of "dense spreading" with nonzero values commonly binary or spherical [20]. The idea of LDS is to use spreading sequences that are the sparse counterparts of the dense spreading sequences of conventional CDMA; a fraction only of the dimensions is filled with nonzero entries [11]. The same concept of LDS can be found in [21] within the framework of time hopping CDMA, where time hopping and chips are mapped to frequency hopping and subbands, respectively. Specifically, the analysis therein can be considered as a reference for LDS in terms of information theoretic bounds.

On the other hand, the massive connectivity of 5G wireless communications is modeled by letting the number of devices to be much larger compared to the number of degrees of freedom, therefore, the asymptotic system behavior should be considered in the large system limit, where the number of users and dimensions go to infinity with same scaling [20]. Note that in CDMA, $N$ is also the number of dimensions. The behavior of DS with random spreading was analyzed in the LSL in pioneering works of Tse and Hanly [58], Tse and Zeitouni [59], Verdú and Shamai [20], and Shamai and Verdú [19]. Subsequently, LDS was similarly analyzed in [21] in the case of a channel without fading. There has been no investigation of the effect of frequency-flat fading so far on the spectral efficiency of LDS.

Therefore, the goal of this chapter is to fill the gap by investigating LDS within the information theoretic framework considered in [19–21] in the presence of frequency-flat fading. We analyze fundamental limits in the LSL when the number of simultaneous transmissions becomes large with respect to the number of degrees of freedom.

### 4.1.2 Other Related Work

Based on the scaling between the number of users and number of degrees of freedom, other related works beyond those mentioned so far investigated either large-scale systems [47–49] or small-scale systems [11–13]. The two different regimes require asymptotic derivations (as the number of users and degrees of freedom grow with same scaling) and non-asymptotic derivations (for finite values of the number of users and degrees of freedom), respectively. The aforementioned literature is detailed as follows:

#### Large-scale system

Most of prior works [48, 49] on LDS in the LSL was derived by means of the replica method, which was first used for DS by Tanaka [60]. Since the replica method is not rigorous, Tanaka's capacity formula was verified (up to a given load, called spinodal, approximately equal to $\beta_s \approx 1.49$) in the LSL by Montanari and Tse in [47], where random spreading with sparse sequences was used in the proof, jointly with belief propagation detection. Adopting the replica method, Raymond and Saad in [49] and Yoshida and Tanaka in [48] analyzed binary sparse CDMA in terms of spectral efficiency with different assumptions on the sparsity level (i.e., the number

of nonzero entries) $N_S$ of signatures (in particular, $N_S$ is a deterministic finite value in [49], whereas $N_S$ is a Poissonian random variable in [48]).

#### Small-scale system

Recent investigations [11–13] analyzed LDS with finite values for the number of users and signal dimensions, in the *overloaded* regime, where the number of users exceeds the number of dimensions. In [11], each user spreads data over a small number of dimensions (e.g., $N_S = 3$) with other dimensions being zero padded. The resulting spreading sequence for each user is then interleaved such that the signature matrix from all $K$ users appears to be very sparse. The analysis focused on the bit error rate for different receiver structures. A comparison with different received powers was also described to address the near-far problem. Using the same framework proposed in [11], an information theoretic analysis of LDS with fading was presented in [12] for a bounded numbers of active users. In particular, the capacity region of time-varying fading LDS channel was analytically determined and tested by simulation, given different sparsity levels and different maximum number of users per dimension.

### 4.1.3 Approach and Contribution

In this chapter, we extend the information theoretic framework of time- and frequency-hopping CDMA considered in [21] for LDS in the presence of frequency-flat fading along the lines of [19]. In [21], the reference channel is the additive white Gaussian noise (AWGN) channel: in order to apply some of the result derived in [21] in an IoT setting, it is mandatory to extend the analysis to channels with fading. We propose an information theoretic analysis where achievable spectral efficiency with different receiver structures is derived for the case of sparse signatures ($N_S = 1$), and compare our results to the spectral efficiency of direct-sequence (DS) CDMA, which represents the archetypal example of dense spreading ($N_S = N$), under the same input constraints such as energy per symbol and bandwidth [19].

The major contributions of this chapter are as follows:

- A rate achievable with linear detection is derived in Theorem 1 in closed form. It is possible to show that sparse signaling outperforms dense signaling when the network is overloaded ($K > N$) and that the effect of fading is to slightly increase the achievable rate in this region.

- The spectral efficiency with optimum detection is derived in Theorem 3 in closed form. It is possible to show that dense signaling outperforms sparse signaling in this setup.

- The spectral efficiency with optimum detection is derived by finding the limiting spectral distribution of a matrix ensemble that jointly describes spreading and fading: this is a mathematical result of independent interest. The combinatorial structure of the moments of such distribution is compared

to the combinatorial interpretations available for the case of LDS and DS without fading.

- The spectral efficiency with optimum detection in the LSL also validates the decoupling principle in the CDMA literature, showing its equivalence to the average rate of a set of parallel channels. Intuitively, the multiuser low-density NOMA with optimum detection may be interpreted as a bank of channels, where each channel experiences an equivalent single-user channel.
- The results provide an insight into the design of signaling in dense networks. As envisioned in the IoT setting, many simple transceivers will be part of large networks: results in this chapter suggest that, in the uplink of such networks, sparse signaling can achieve a rate several times larger than that achievable via dense signaling.

### 4.1.4 Organization

The chapter is organized as follows. Section 4.2 introduces the reference model for LDS based on the general framework of traditional DS with the same energy and bandwidth constraints. The most important results in the literature relevant to our analysis are recalled in Section 4.3. Achievable spectral efficiency of LDS with linear and optimum receivers are presented in Section 4.4. Finally, conclusions based on the comparison of fundamental limits of LDS in 5G network are drawn in Section 4.6.

## 4.2 Reference model

The proposed reference model of a LDS system in the presence of frequency-flat fading follows the traditional discrete complex-valued CDMA model

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{b} + \boldsymbol{n}, \tag{4.1}$$

where: $\boldsymbol{y} \in \mathbb{C}^N$ is the received signal; $\boldsymbol{b} = [b_1, \ldots, b_K]^\mathsf{T} \in \mathbb{C}^K$ is the vector of symbols transmitted by the $K$ users; $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K] \in \mathbb{R}^{N\times K}$ is a random spreading matrix, column $i$ being the unit-norm spreading sequence of user $i$; $\boldsymbol{A} \in \mathbb{C}^{K\times K}$ is a diagonal matrix of complex-valued fading coefficients $\mathrm{diag}(a_1, \ldots, a_K)$; and $\boldsymbol{n} \in \mathbb{C}^N$ is a circularly symmetric Gaussian vector with a zero mean and covariance $\mathcal{N}_0\boldsymbol{I}$. Users transmit independent symbols and obey the power constraint $\mathbb{E}[|b_k|^2] \leqslant \mathcal{E}$ for all $k$, hence

$$\mathbb{E}[\boldsymbol{b}\boldsymbol{b}^*] = \mathcal{E}\boldsymbol{I}. \tag{4.2}$$

The load of the system is defined as the ratio between the number of users $K$ and the number of dimensions $N$, and is denoted by $\beta := K/N$. Systems with $\beta < 1$ and $\beta > 1$ are referred to as underloaded and overloaded systems, respectively.

Both LDS and DS systems can be modeled by (5.2) with sparse and dense spreading matrix $\boldsymbol{S}$, respectively. In the simplest models, all elements of $\boldsymbol{S}$ are nonzero in DS, e.g. $s_{ki} \in \{\pm 1/\sqrt{N}\}$, while all but one element per column is nonzero in LDS, i.e. $\boldsymbol{s}_k \in \{\pm \boldsymbol{e}_i^N\}_{i=1,\ldots,N}$. For the sake of clarity, we define rigorously below what we mean by sparse vector and sparse matrix.

*Definition* 1 (Sparse vector). A vector $\boldsymbol{v} \in \mathbb{R}^N$ is $N_S$-sparse if the cardinality of the set of its nonzero elements is $N_S$, i.e. $\|\boldsymbol{v}\|_0 := |\{v_i \neq 0\}_{i=1,\ldots,N}| = N_S$.

*Definition* 2 (Sparse matrix). A matrix $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K]$ is $N_S$-sparse if each column $\boldsymbol{s}_k$ is an $N_S$-sparse vector.

A reference model for time- and frequency-hopping CDMA was presented in [21] building on the seminal paper [20]. The present work extends the model of [21] by introducing fading along the lines of [19]. Notice that the assumption underpinning the fading model is that fading coefficients do not change over the whole signature, and more generally over the whole coherence block. This assumption may seem to clash with the pursued large system analysis since the latter requires increasingly large signatures. However, notice that the LSL is only used to derive closed form expressions of performance of interest: It is well known that results derived in the LSL are in fact very good approximations of performance of finite systems. The only important assumption is to keep the same load $\beta$ in the finite system and in the large system.

In the following, we consider the very sparse scenario corresponding to sparse matrices with 1-sparse column vectors. In this case, each spreading sequence $\boldsymbol{s}_k$ contains only one nonzero element, equal to either +1 or −1, with equal probability. Hence, the energy of the sequence is concentrated in just one nonzero pulse, while in DS, the energy is uniformly spread over all $N$ dimensions.

System performance is measured by spectral efficiency $C$, defined as the total number of bits per dimension, that can be reliably transmitted [19–21]. The per-symbol signal-to-noise ratio (SNR) is given by [61]

$$\gamma := \frac{\frac{1}{K}\mathbb{E}[\,\|\boldsymbol{b}\|^2]}{\frac{1}{N}\mathbb{E}[\,\|\boldsymbol{n}\|^2]} = \frac{N}{K} \cdot \frac{b}{N} \cdot \frac{\mathcal{E}_\mathrm{b}}{\mathcal{N}_0} = \frac{1}{\beta} \cdot C \cdot \eta, \tag{4.3}$$

where $C = b/N$ is expressed in bits per dimension, $b$ is the number of bits encoded in $\boldsymbol{b}$, $\mathbb{E}[\,\|\boldsymbol{b}\|^2] = b\mathcal{E}_\mathrm{b}$, $\mathbb{E}[\,\|\boldsymbol{n}\|^2] = N\mathcal{N}_0$, and $\eta := \mathcal{E}_\mathrm{b}/\mathcal{N}_0$.

## 4.3 Previous Results

In this section, we summarize the results in the literature that are most relevant to our analysis, namely spectral efficiency for LDS without fading and spectral efficiency of DS with and without fading.

### Spectral efficiency in the absence of fading for LDS and DS

The model in (5.2) reduces to that in [21] when $\boldsymbol{A} = \boldsymbol{I}$ (no fading). Optimum decoding with LDS achieves the following spectral efficiency:

$$C_\mathrm{lds}^\mathrm{opt}(\beta, \gamma) = \sum_{k \geqslant 0} \frac{\beta^k e^{-\beta}}{k!} \log_2(1 + k\gamma) \text{ bits/s/Hz}. \tag{4.4}$$

Spectral efficiency with DS is [19, 20]

$$C_{\text{ds}}^{\text{opt}}(\beta,\gamma) = \beta \log_2\left(1+\gamma-\frac{1}{4}\mathcal{F}(\gamma,\beta)\right) + \log_2\left(1+\beta\gamma-\frac{1}{4}\mathcal{F}(\gamma,\beta)\right) - \frac{1}{4\log 2}\cdot\frac{\mathcal{F}(\gamma,\beta)}{\gamma} \text{ bits/s/Hz}, \quad (4.5)$$

where

$$\mathcal{F}(x,z) = \left(\sqrt{x(1+\sqrt{z})^2+1}-\sqrt{x(1-\sqrt{z})^2+1}\right)^2. \quad (4.6)$$

Linear detectors, such as single-user matched filter (SUMF), zero-forcing (ZF), and minimum mean square error (MMSE), result in the same mutual information with LDS. An achievable spectral efficiency for these multiple access channels is $R_{\text{lds}}^{\text{sumf}} = \beta I(b_1; r_1|\boldsymbol{S})$ (b/s/Hz) with $\boldsymbol{b}$ Gaussian and $\boldsymbol{S}$ sparse, where $I(b_1; r_1|\boldsymbol{S})$ is the achievable rate (bits/symbol) of user 1:

$$R_{\text{lds}}^{\text{sumf}}(\beta,\gamma) = R_{\text{lds}}^{\text{zf}}(\beta,\gamma) = R_{\text{lds}}^{\text{mmse}}(\beta,\gamma) = \beta\sum_{k\geqslant 0}\frac{\beta^k e^{-\beta}}{k!}\log_2\left(1+\frac{\gamma}{k\gamma+1}\right) \text{ bits/s/Hz}. \quad (4.7)$$

Differently from LDS, linear detectors with DS achieve different spectral efficiency. Among the above mentioned linear detectors, MMSE achieves the highest spectral efficiency, which is equal to [20]

$$C_{\text{ds}}^{\text{mmse}}(\beta,\gamma) = \beta\log\left(1+\gamma-\frac{1}{4}\mathcal{F}(\gamma,\beta)\right). \quad (4.8)$$

**Spectral efficiency in the presence of fading for DS**

In the presence of fading, spectral efficiency with optimum decoding is [19]

$$C_{\text{ds}}^{\text{opt}}(\beta,\gamma) = C_{\text{ds}}^{\text{mmse}}(\beta,\gamma) + \frac{\eta-1-\log\eta}{\log 2} \quad (4.9)$$

where $\eta > 0$ satisfies the following fixed point equation

$$\eta = 1-\beta+\beta\,\mathbb{E}\left[\frac{1}{1+\eta|a|^2\gamma}\right], \quad (4.10)$$

and $C_{\text{ds}}^{\text{mmse}}(\beta,\gamma)$ is spectral efficiency with MMSE, given by

$$C_{\text{ds}}^{\text{mmse}}(\beta,\gamma) = \beta\,\mathbb{E}[\log_2(1+\gamma|a|^2\eta)]. \quad (4.11)$$

It is noteworthy that fading increases spectral efficiency with MMSE at high load. The intuition provided in [19, Section III-C] is that some user appears very low-powered at the receiver, thus the "interference population," i.e., the number of effective interferers is reduced. A similar behavior is not observed with ZF, which removes all interference irrespective of power. This effect is called "interference population control."

## 4.4 Spectral Efficiency of LDS with Frequency-Flat Fading

In this section, we derive spectral efficiency with a bank of single-user matched filters and independent decoding in Section 4.4.1 and with optimum decoding in Section 4.4.2.

### 4.4.1 Single-User Matched Filter (SUMF)

The decision variable for user 1 is

$$\begin{aligned} r_1 &= \boldsymbol{s}_1^{\mathsf{T}} \boldsymbol{y} \\ &= \boldsymbol{s}_1^{\mathsf{T}} \left( \sum_{k=1}^{K} \boldsymbol{s}_k a_k b_k \right) + \boldsymbol{s}_1^{\mathsf{T}} \boldsymbol{n} \\ &= a_1 b_1 + \sum_{k=2}^{K} \boldsymbol{s}_1^{\mathsf{T}} \boldsymbol{s}_k a_k b_k + \boldsymbol{s}_1^{\mathsf{T}} \boldsymbol{n}, \end{aligned} \tag{4.12}$$

where the last step follows from the signatures being unit norm. Assuming Gaussian coding, $b_k \sim \mathcal{N}_{\mathbb{C}}(0, \mathcal{E})$, the conditional mutual information (bits/symbol) for user 1 is

$$\begin{aligned} I(r_1; b_1 | \boldsymbol{S}, \boldsymbol{A}) &= I(y_1; b_1 | \rho_{12}, \ldots, \rho_{1K}, a_1, \ldots, a_K) \\ &= \mathbb{E}\left[ \log_2 \left( 1 + \frac{|a_1|^2 \gamma}{1 + \gamma \sum_{k=2}^{K} \rho_{1k}^2 |a_k|^2} \right) \right], \end{aligned} \tag{4.13}$$

where $\rho_{1k} := \boldsymbol{s}_1^{\mathsf{T}} \boldsymbol{s}_k$ and the expectation is taken with respect to $\{\rho_{12}, \ldots, \rho_{1K}\}$ and $\{a_1, \ldots, a_K\}$. The corresponding mutual information of the multiuser channel is

$$R_{\text{lds}}^{\text{sumf}}(\beta, \gamma) := \beta I(r_1; b_1 | \boldsymbol{S}, \boldsymbol{A}) \text{ bits/s/Hz}.$$

In the following theorem we propose an explicit form of (4.13) for 1-sparse matrices (cf. Definition 2).

*Theorem* 1. Let $\boldsymbol{S} \in \mathbb{R}^{N \times K}$ be a 1-sparse spreading matrix. In the LSL, the following rate is achievable with a bank of SUMF detectors:

$$R_{\text{lds}}^{\text{sumf}}(\beta, \gamma) = \frac{\beta}{\log 2} \int_0^1 \frac{e^{-t\left(\beta + \frac{1}{1-t} \cdot \frac{1}{\gamma}\right)}}{1-t} \, dt \text{ bits/s/Hz}. \tag{4.14}$$

*Proof* See Appendix 4.A.

The result in Theorem 1 allows us to study asymptotics for low and high SNR. In the low-SNR regime, the minimum energy per bit per noise level is given by (see Appendix 4.B for the proof)

$$\eta_{\min} = \lim_{\gamma \to 0} \frac{\beta \gamma}{R_{\text{lds}}^{\text{sumf}}(\beta, \gamma)} = \log 2 \text{ dB}, \tag{4.15}$$

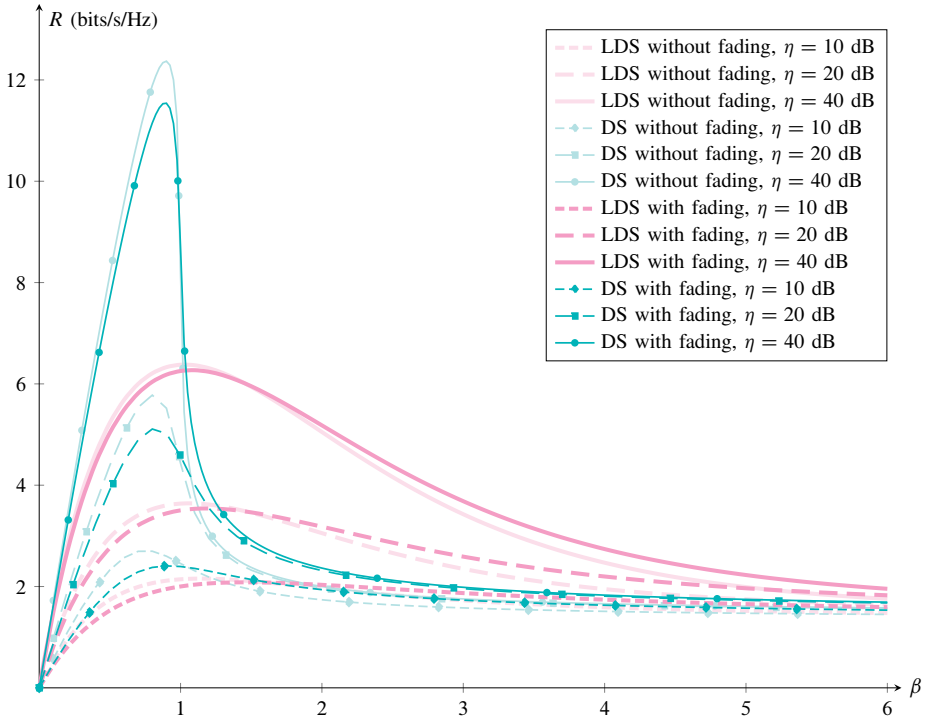



Figure 4.1. Achievable spectral efficiency (bits/s/Hz) of LDS with SUMF detection (thick lines) and DS with MMSE detection (thin lines with marks) for several values of $\eta$ as a function of the load in the presence (dark shade) or absence (light shade) of fading.

as in the case without fading, and with fading and dense spreading. Note that (4.15) holds for any $\beta > 0$. The slope at $\eta = \eta_{\min}$ is (see Appendix 4.C for the proof)

$$\mathcal{S}_0^{\text{sumf}} = 2\log 2 \lim_{\gamma\to 0} \frac{\left(\frac{\partial}{\partial\gamma} R_{\text{lds}}^{\text{sumf}}\right)^2}{-\frac{\partial^2}{\partial\gamma^2} R_{\text{lds}}^{\text{sumf}}} = \frac{\beta}{1+\beta} \text{ bits/s/Hz/(3 dB)}, \tag{4.16}$$

that is the same slope achieved with dense signaling. In the high-SNR regime, rate grows logarithmically with high-SNR slope equal to (see Appendix 4.D for the proof)

$$\mathcal{S}_\infty^{\text{sumf}} = \log 2 \lim_{\gamma\to\infty} \gamma \frac{\partial R_{\text{lds}}^{\text{sumf}}}{\partial\gamma} = \beta e^{-\beta} \text{ bits/s/Hz/(3 dB)}, \tag{4.17}$$

which is the same as LDS without fading, and, compared with the high-SNR slope achieved by DS with MMSE,

$$\mathcal{S}_{\infty,\text{ds}}^{\text{mmse}} = \beta \mathbb{1}_{\{\beta\in[0,1)\}} + \frac{1}{2}\mathbb{1}_{\{\beta=1\}} + 0\mathbb{1}_{\{\beta>1\}}, \tag{4.18}$$

shows that, for $\beta > 1$, LDS is preferable to DS.

Figure 4.1 shows the achievable spectral efficiency with linear detection, and compares DS and LDS in the presence and absence of fading. In the presence of fading, the same qualitative phenomenon observed without fading holds, namely

LDS outperforms DS, when load is approximately higher than unity. We stress that the curves for DS are capacities whereas the curves for LDS are merely achievable rates, and that this is sufficient to claim that LDS outperforms DS in the overloaded regime. The gap in performance with and without fading follows the same pattern for both DS and LDS, namely rates are decreased in the underloaded regime and increased in the overloaded regime. Finally, we notice that both LDS and DS are characterized by the same slope at $\beta = 0$ and the same asymptotic value as $\beta \to \infty$.

### 4.4.2 Optimum decoding

The spectral efficiency achieved with optimum decoding is the maximum (over the distributions on $\boldsymbol{b}$) normalized mutual information between $\boldsymbol{b}$ and $\boldsymbol{y}$ knowing $\boldsymbol{S}$ and $\boldsymbol{A}$, which is given by [19, 62]

$$C_N^{\text{opt}}(\beta,\gamma) = \frac{1}{N}\log_2 \det(\boldsymbol{I} + \gamma \boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*). \tag{4.19}$$

We can express (4.19) in terms of the set of eigenvalues of the Gram matrix $\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*$, $\{\lambda_n(\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*) : 1 \leqslant n \leqslant N\}$, as follows:

$$C_N^{\text{opt}}(\beta,\gamma) = \int_0^\infty \log_2(1+\gamma\lambda)\, dF_N^{\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*}(\lambda), \tag{4.20}$$

being $F_N^{\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*}(x)$ the *empirical spectral distribution* (ESD) of $\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*$, namely [63,64]:

$$F_N^{\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*}(x) := \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{\lambda_n(\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*)\leqslant x\}}. \tag{4.21}$$

Being $\boldsymbol{S}$ and $\boldsymbol{A}$ random, also $F_N^{\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*}$ is random. In the LSL, as is well known, the ESD can admit a limit (in probability or stronger sense), which is called *limiting spectral distribution* (LSD) [64] and is denoted by $F(x)$. Hence, if the limit exists, spectral efficiency $C_N^{\text{opt}}(\gamma)$ converges to

$$C^{\text{opt}}(\beta,\gamma) = \int_0^\infty \log_2(1+\gamma\lambda)\, dF(\lambda)\,. \tag{4.22}$$

Our main goal is, therefore, to find the LSD of the matrix ensemble $\{\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*\}$. To this end, we compute in Theorem 2 the average moments of the ESD in the LSL and prove convergence in probability of the sequence of (random) moments of the ESD,

$$m_L := \frac{1}{N}\text{tr}(\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*)^L = \int_0^\infty \lambda^L\, dF_N^{\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*}(\lambda), \tag{4.23}$$

to the (nonrandom) moments of the LSL. Then, by verifying Carleman's condition, Lemma 1 shows that these moments uniquely specify the LSD [65]. Finally, we use the LSD to derive the spectral efficiency in the LSL in Theorem 3.

*Theorem* 2. Given the matrix ensemble $\{\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*\}$ with $\boldsymbol{S}$ an $N \times K$ sparse spreading matrix and $\boldsymbol{A}$ a $K \times K$ diagonal matrix of Rayleigh fading coefficients, it results

$$m_L \xrightarrow{\text{P}} \bar{m}_L := \sum_{l=1}^{L} \left\lfloor {L \atop l} \right\rfloor \beta^l, \tag{4.24}$$

where $\left\lfloor {L \atop l} \right\rfloor := \binom{L-1}{l-1}\frac{L!}{l!}$ denotes the *Lah numbers* [66].

Table 4.1. Summary of LSDs, moments, and their combinatorial structure for different scenarios with and without fading.

| **Scenario** | DS with no fading | LDS with no fading | LDS with fading |
|---|---|---|---|
| **LSD** | Marčenko-Pastur law | Poisson law | Compound Poisson law |
| **LSD moment** $\bar{m}_L$ | $\sum_{l=1}^{L} \mathcal{N}_l \beta^l$ | $\sum_{l=1}^{L} \left\{ {L \atop l} \right\} \beta^l$ | $\sum_{l=1}^{L} \lfloor {L \atop l} \rfloor \beta^l$ |
| **Coefficient** | $\mathcal{N}_l = \frac{1}{L}\binom{L}{l}\binom{L}{l-1}$ | $\left\{ {L \atop l} \right\} = \frac{1}{l!}\sum_{j=0}^{l}(-1)^{l-j}\binom{l}{j} j^L$ | $\lfloor {L \atop l} \rfloor = \binom{L-1}{l-1}\frac{L!}{l!}$ |

*Proof.* See Appendix 4.E. □

In particular, $\bar{m}_L$ is the $L^{\text{th}}$ moment of the random variable $\sum_{j=1}^{J} Z_j$ where $J$ is distributed according to a Poisson law with mean $\beta$ and, conditionally on $J$, $\{Z_j : 1 \leqslant j \leqslant J\}$ is a set of i.i.d. exponentially distributed random variables with unit rate.

In the following lemma, we verify that the LSD is uniquely determined by the sequence of moments $(\bar{m}_L)_{L \geqslant 1}$.

*Lemma* 1. The sequence of moments $(\bar{m}_L)_{L \geqslant 1}$ satisfies the Carleman's condition, namely the series $\sum_{k \geqslant 1} \bar{m}_{2k}^{-1/(2k)}$ diverges.

*Proof.* See Appendix 4.F. □

Therefore, Theorem 2 and Lemma 1 imply that the probability measure $F(\lambda)$ in (4.22) is the probability measure of a compound Poisson distribution generated by the sum of a mean-$\beta$ Poissonian number of unit-rate exponentially distributed random variables:

$$F(d\lambda) = e^{-\beta}\delta_0(d\lambda) + \sum_{k \geqslant 1} \frac{e^{-\beta}\beta^k}{k!} \cdot \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!}\, d\lambda\,. \tag{4.25}$$

The spectral efficiency in the LSL is thus given by the average rate experienced through a set of parallel channels, indexed by $k = 1, 2, \cdots$, with signal-to-noise ratio equal to $\lambda\gamma$, used with probability $(e^{-\beta}\beta^k/k!) \cdot (e^{-\lambda}\lambda^{k-1}/(k-1)!)\, d\lambda$. Indeed, this observation may validate the claim by Guo and Verdú that in the LSL, the CDMA channel followed by multiuser detection can be decoupled into a bank of parallel Gaussian channels, each channel per user [67]. This is referred to as decoupling principle, which leads to the convergence of the mutual information of multiuser detection for each user to that of equivalent single-user Gaussian channel as the number of users go to infinitive, given the same input constraints. Given that the randomness of ESD vanishes in the LSL (cf. Theorem 2), one may invoke the "self-averaging" property in the statistical physics [67]. Similarly to CDMA, the self-averaging principle yields to the strong property that for almost all realizations of the spreading sequences and noise of low-density NOMA, the macroscopic quantity

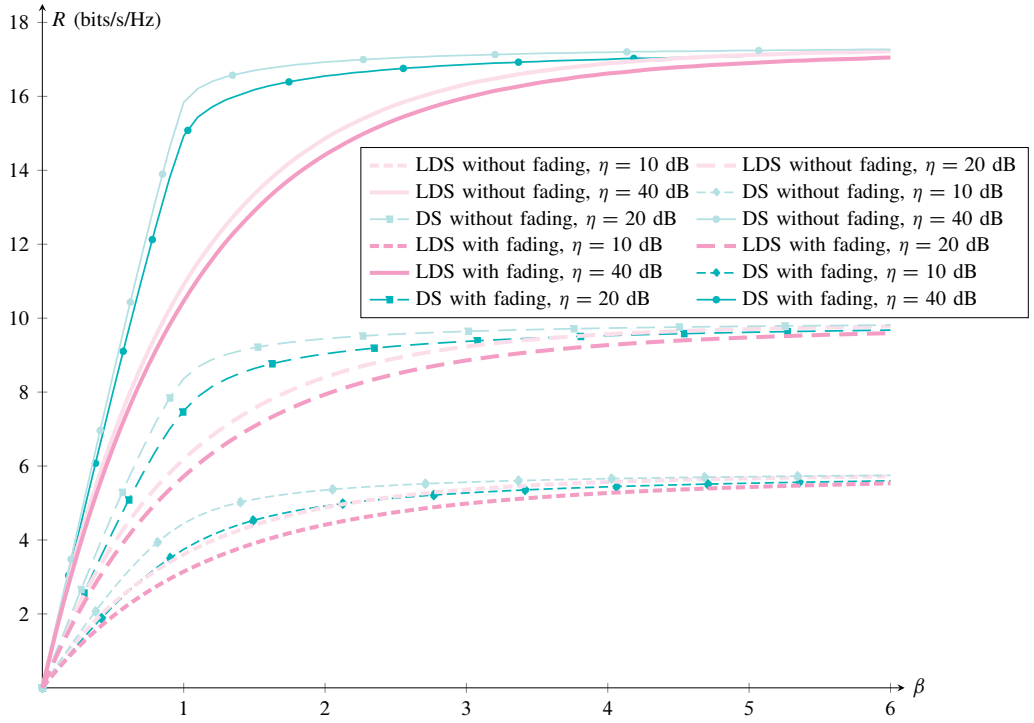



Figure 4.2. Achievable spectral efficiency (bits/s/Hz) of LDS (thick lines) and DS (thin lines with marks) with optimum decoding for several values of $\eta$ as a function of the load in the presence (dark shade) or absence (light shade) of fading.

(spectral efficiency in this case) converges to an equivalent deterministic quantity in the large system regime.

*Theorem* 3. The spectral efficiency with optimum decoding in the LSL is given by

$$C^{\text{opt}}(\beta,\gamma) = \sum_{k\geqslant 1} \frac{e^{-\beta}\beta^k}{k!} \int_0^\infty \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} \log_2(1+\gamma\lambda)\, d\lambda\,. \tag{4.26}$$

*Proof.* Plug (4.25) into (4.22) and commute summation and integration, which is followed from Tonelli's theorem. □

Similarly to the previous section, it is interesting also here to study the asymptotic behavior of spectral efficiency as a function of $\eta$. In the low-SNR regime, the minimum energy per bit per noise level is given by (see Appendix 4.G for the proof)

$$\eta_{\min} = \lim_{\gamma\to 0} \frac{\beta\gamma}{C^{\text{opt}}_{\text{lds}}(\beta,\gamma)} = \log 2 \text{ dB}, \tag{4.27}$$

as in the case without fading, and with fading and dense spreading, irrespective of $\beta > 0$. The slope at $\eta = \eta_{\min}$ is (see Appendix 4.H for the proof)

$$\mathcal{S}_0^{\text{opt}} = 2\log 2 \lim_{\gamma\to 0} \frac{(\frac{\partial}{\partial\gamma} C^{\text{opt}}_{\text{lds}})^2}{-\frac{\partial^2}{\partial\gamma^2} C^{\text{opt}}_{\text{lds}}} = \frac{2\beta}{\beta+2} \text{ bits/s/Hz/(3 dB)}, \tag{4.28}$$

which is the same as with dense signaling in the presence of fading (cf. (147) in [19]). In the high-SNR regime, rate grows logarithmically with high-SNR slope equal to (see Appendix 4.I for the proof)

$$\mathcal{S}_{\infty}^{\text{opt}} = \log 2 \lim_{\gamma\to\infty} \gamma \frac{\partial C_{\text{lds}}^{\text{opt}}}{\partial \gamma} = 1 - e^{-\beta} \text{ bits/s/Hz/(3 dB)}, \tag{4.29}$$

which is the same as without fading.

Figure 4.2 shows the achievable spectral efficiency with optimum decoding and compares DS and LDS, in the presence and absence of fading. It is shown that, in general, LDS underperforms DS irrespective of fading; however, the main gap is concentrated around $\beta = 1$, and decreases as load goes either to 0 or $\infty$.

We conclude this section by highlighting the combinatorial connection between the moments found in Theorem 2 and moments of the Marčenko-Pastur and Poisson laws (see also Table 4.1), which correspond to the limiting spectral distributions of dense [20] and sparse [21] schemes, respectively. We showed that the $L^{\text{th}}$ moment is essentially a polynomial in $\beta$ with coefficients equal to Lah numbers. Similar results hold for dense and sparse schemes without fading, where Lah numbers are replaced by Narayana numbers and Stirling numbers of the second kind, respectively. All numbers are well-known in combinatorics: Narayana numbers enumerate non-crossing partitions into nonempty subsets; Stirling numbers of the second kind enumerate partitions into nonempty subsets; and Lah numbers enumerate partitions into nonempty linearly ordered subsets [66].

## 4.5 Synopsis of results for LDS vs DS systems

We collect the main results on LDS and DS systems from another perspective, namely for the case of fixed load and variable $\eta$, in Figs. 4.3 and 4.4. For fixed $\beta$, one can find the spectral efficiency as a function of $\eta$ by solving (4.3) with respect to $\gamma$ and computing the spectral efficiency for such value of $\gamma$. Achievable spectral efficiency (b/s/Hz), as a function of $\eta$ with optimum and linear detection in the presence and absence of fading, is shown for $\beta = 1$ and $\beta = 2$, respectively. To summarize the sources, results for DS were derived in [20] without fading and in [19] with fading, whereas results for LDS without fading were derived in [21].

Both figures show that all schemes are equivalent in the low-SNR regime, and that DS outperforms LDS with optimum decoding, particularly in the high-SNR regime, where spectral efficiency of the two schemes is characterized by different slopes. With linear detection, the scenario is completely different: when load increases beyond approximately unity, LDS outperforms DS, with a widening gap as $\eta$ increases. Indeed, LDS keeps a positive high-SNR slope while DS cannot afford it (cf. (4.17) vs. (4.18)). Note the effect of the "interference population control" with DS (cf. Section 4.3) on both figures: spectral efficiency with fading is higher than spectral efficiency without fading. A similar behavior holds with LDS as shown on Fig. 4.4.

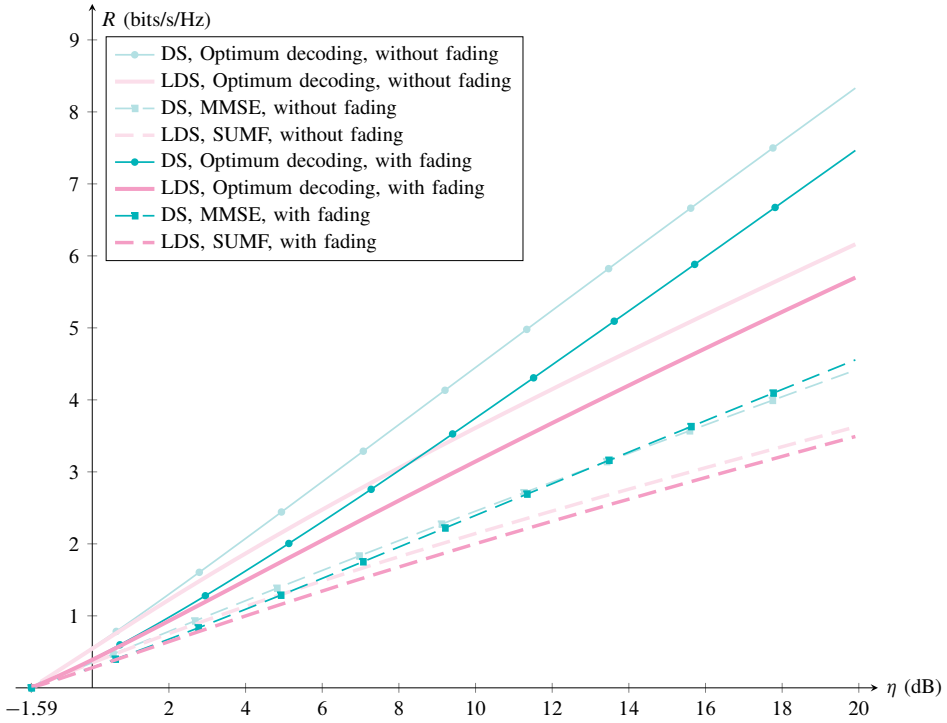



Figure 4.3. Achievable spectral efficiency (bits/s/Hz) of LDS (thick lines) and DS (thin lines with marks) with optimum detection as a function of $\eta$ (dB) with load $\beta = 1$ in the presence (dark shade) or absence (light shade) of fading.

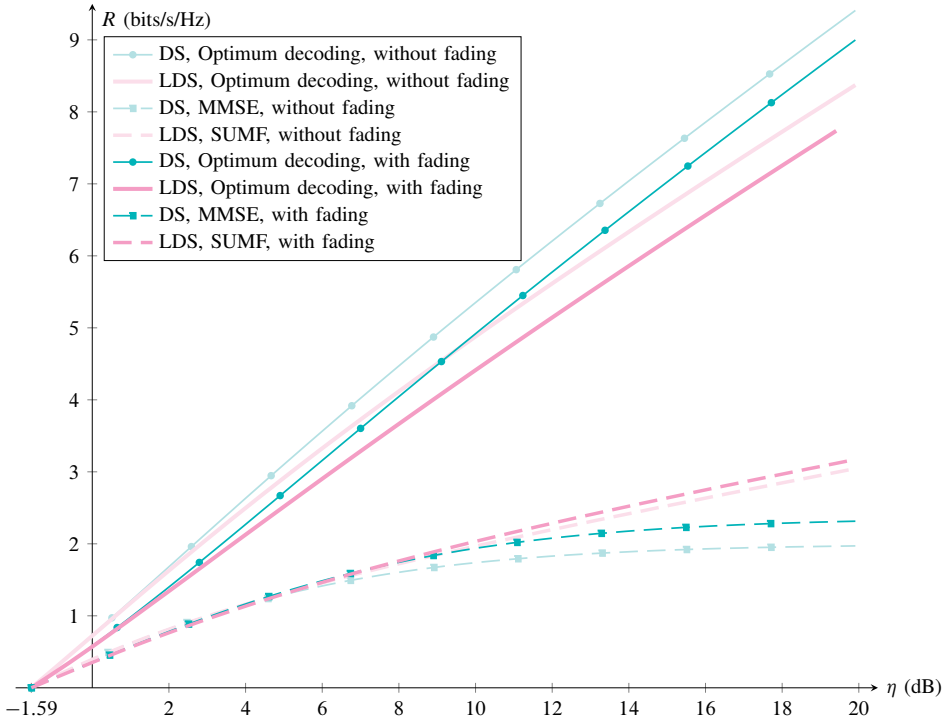



Figure 4.4. Achievable spectral efficiency (bits/s/Hz) of LDS (thick lines) and DS (thin lines with marks) with optimum detection as a function of $\eta$ (dB) with load $\beta = 2$ in the presence (dark shade) or absence (light shade) of fading.

## 4.6 Conclusion

In this chapter, a theoretical analysis of LDS systems in the presence of flat fading in terms of spectral efficiency with linear and optimum receivers was carried out in the LSL, i.e., as both the number of users $K$ and the number of degrees of freedom $N$ grow unboundedly, with a finite ratio $\beta = K/N$. Spectral efficiency was derived as a function of the load $\beta$ and signal-to-noise ratio $\gamma$. The framework used extended the model in [21], which was build on the seminal work [20], to the case with fading, along the lines of [19] in both the underloaded and overloaded regimes corresponding to $\beta < 1$ and $\beta > 1$, respectively. The behavior of LDS was compared against DS. Since spreading sequences of LDS are indeed the sparse counterparts of DS ones, the reference model introduced in [21] for time hopping CDMA could also be adopted in the present context, in order to describe both DS and LDS, where the spreading matrix $\mathbf{S} \in \mathbb{C}^{N \times K}$ of the model defines the sparsity of the system. Note that the model in [21] originally defined for time hopping CDMA, also fits the dual frequency domain in describing frequency-hopping CDMA.

In the absence of fading, previous work showed that, in the LSL, DS has higher spectral efficiency than LDS when the system is underloaded ($\beta < 1$). However, a drastic drop occurs at about $\beta = 1$, and eventually, in the overloaded regime ($\beta > 1$), LDS outperforms DS [21]. In this chapter, we were to able to show that this is the case also in the presence of fading.

This is particularly important in view of massive deployment of wireless devices and ultra-densification of the network towards 5G. Overloaded systems, where the number of resources is lower than the number of users accessing the network, will play a pivotal role in 5G, and this work provides a theoretical ground for choosing LDS with respect to DS, and more generally choosing sparsity over density in signaling formats. In the presence of flat fading, [19] showed that in the overloaded condition, the fading effect, for DS, is to enhance spectral efficiency. In this chapter, we filled the gap by showing that, in the LSL, LDS with fading achieved higher spectral efficiency than DS in the overloaded mode. Moreover, the dominance of LSD-CDMA over DS in terms of spectral efficiency performance than DS goes stronger when $E_b/N_0$ increases.

In conclusion, results of this chapter suggest that for overloaded systems, it is suggested to adopt low-density spreading with linear receivers. This property holds despite fading. Moreover, low-density spreading CDMA, employing linear receivers, achieves even higher spectral efficiency with respect to the dense system in the presence of flat fading, which is a practical issue in wireless communications.

Future investigations will focus on refining the understanding of overloaded systems with a more general structure of the sparsity of spreading sequences, e.g., when $N_S > 1$. It is interesting to understand which value of $N_S$ represents the boundary between dense and low-dense systems, in terms of capacity, and more generally how the system behaves as a function of $N_S$.

# Appendices

## 4.A Proof of Theorem 1

In order to compute (4.13) we need to find the distribution of $\zeta := \sum_{k=2}^{K} \rho_{1k}^2 |a_k|^2$. We recall that $\rho_{1k} := \boldsymbol{s}_1^{\mathsf{T}} \boldsymbol{s}_k$, hence the moment generating function (MGF) of $\rho_{1k}^2$ is

$$M_{\rho^2}(t) := \mathbb{E}[e^{t\rho^2}] = \left(1 - \frac{1}{N}\right) + \frac{1}{N} e^t,$$

irrespective of $k$. Since $|a_k|^2$ is exponentially distributed with unit rate, we conclude

$$M_{\rho^2|a|^2}(t) := \mathbb{E}[e^{t\rho^2|a|^2}] = \mathbb{E}[M_{\rho^2}(t|a|^2)] = 1 + \frac{t}{(1-t)N}.$$

Therefore, the MGF of $\zeta$ is $(M_{\rho^2|a|^2}(t))^{K-1}$, and in the LSL

$$M_\zeta(t) \to e^{\beta \frac{t}{1-t}}.$$

Now, we express the logarithm via the following integral representation [68]

$$\log(1+x) = \int_0^\infty \frac{1}{s}(1 - e^{-sx}) e^{-s}\, ds,$$

which is valid for $x \geqslant 0$. Hence one has

$$\log\left(1 + \frac{|a|^2}{\zeta + 1/\gamma}\right) = \int_0^\infty \frac{ds}{s} e^{-s/\gamma}(1 - e^{-s|a|^2}) e^{-s\zeta},$$

and by taking the expectation and changing variable, $t = \frac{s}{1+s}$,

$$\mathbb{E}\left[\log\left(1 + \frac{|a|^2}{\zeta + 1/\gamma}\right)\right] = \int_0^1 \frac{e^{-t\left(\beta + \frac{1}{(1-t)\gamma}\right)}}{1-t} dt.$$

## 4.B Proof of (4.15)

With a change of variable, we can rewrite $R_{\text{lds}}^{\text{sumf}}(\beta, \gamma)$ as follows:

$$R_{\text{lds}}^{\text{sumf}}(\beta, \gamma) = \frac{\beta}{\log 2} g_\beta(1/\gamma),$$

where

$$g_\beta(\alpha) := \int_0^\infty dz\, e^{-\alpha z} e^{-\beta \frac{z}{1+z}} \frac{1}{1+z}. \tag{.30}$$

We observe that $\eta_{\min}$ is expressed in terms of $g_\beta(\alpha)$ as follows:

$$\eta_{\min} = \lim_{\alpha \to \infty} \frac{\log 2}{\alpha g_\beta(\alpha)},$$

hence we need to study $\alpha g_\beta(\alpha)$ as $\alpha \to \infty$. Since $g_\beta(\alpha)$ does not admit a closed form, we have to study the specific integral in (.30). The basic observation is that the term $e^{-\beta \frac{z}{1+z}}$ is bounded on the integration interval from below and above, namely

$e^{-\beta\frac{z}{1+z}} \in (e^{-\beta}, 1]$. Furthermore, most of the mass is concentrated in a neighborhood of $z = 0$, as $\alpha$ increases. It makes sense to partition the domain $[0, \infty)$ in two subintervals, $[0, \epsilon)$ and $[\epsilon, \infty)$, for some $\epsilon > 0$ fixed:

$$g_\beta(\alpha) = \int_0^\epsilon dz\, e^{-\alpha z} e^{-\beta\frac{z}{1+z}} \frac{1}{1+z} + \int_\epsilon^\infty dz\, e^{-\alpha z} e^{-\beta\frac{z}{1+z}} \frac{1}{1+z}. \tag{.31}$$

The first integral in (.31) is upper and lower bounded by

$$C_1(\epsilon) \int_0^\epsilon dz\, e^{-\alpha z} \frac{1}{1+z}, \tag{.32}$$

with $C_1(\epsilon) = \bar{C}_1(\epsilon) := 1$ and $C_1(\epsilon) = \underline{C}_1(\epsilon) := e^{-\beta\frac{\epsilon}{1+\epsilon}}$, respectively. Similarly, the second integral in (.31) is upper and lower bounded by

$$C_2(\epsilon) \int_\epsilon^\infty dz\, e^{-\alpha z} \frac{1}{1+z}, \tag{.33}$$

with $C_2(\epsilon) = \bar{C}_2(\epsilon) := \underline{C}_1(\epsilon)$ and $C_2(\epsilon) = \underline{C}_2(\epsilon) := e^{-\beta}$, respectively. The integrals in (.32)–(.33) can be expressed by means of known functions,

$$\begin{aligned} \alpha \int_0^\epsilon dz\, e^{-\alpha z} \frac{1}{1+z} &= \alpha e^\alpha [E_1(\alpha) - E_1(\alpha(1+\epsilon))] \\ &= 1 + O(1/\alpha), \end{aligned}$$

where $E_1(x)$ denotes the exponential integral[1] for $x > 0$, for which the following asymptotic expansion holds $\alpha e^{-\alpha} E_1(\alpha) = 1 - \alpha^{-1} + O(\alpha^{-2})$, and

$$\begin{aligned} \alpha \int_\epsilon^\infty dz\, e^{-\alpha z} \frac{1}{1+z} &= \alpha e^\alpha E_1(\alpha(1+\epsilon)) \\ &\leqslant e^{-\epsilon\alpha}(1+\epsilon)^{-1}, \end{aligned}$$

which vanishes as $\alpha \to \infty$, where the inequality follows from the standard bracketing of $E_1$ through elementary functions. Therefore, we proved that, for all $\epsilon > 0$, the term

$$C_1(\epsilon)\ \alpha \int_0^\epsilon dz\, e^{-\alpha z} \frac{1}{1+z} = C_1(\epsilon) + O(1/\alpha), \tag{.34}$$

contributes finitely to the integral, while the term

$$C_2(\epsilon)\ \alpha \int_\epsilon^\infty dz\, e^{-\alpha z} \frac{1}{1+z} \to 0, \tag{.35}$$

asymptotically vanishes. Hence, $\alpha g_\beta(\alpha) \to C_1(\epsilon)$ as $\alpha \to \infty$, and the result follows since $\epsilon > 0$ is arbitrary.

[1] $E_n(x) := \int_1^\infty dt\, \frac{1}{t^n} e^{-xt}$ for all $x > 0$ and $n$ positive integer.

## 4.C Proof of (4.16)

A sketch of the proof is provided. The slope can be written as follows:

$$\mathcal{S}_0 = \beta \lim_{\alpha\to\infty} \frac{\alpha I_1(\alpha)^2}{I_1(\alpha) - \frac{\alpha}{2} I_2(\alpha)}, \tag{.36}$$

where

$$I_k(\alpha) := \int_0^\infty dz\, \frac{z^k}{1+z} e^{-z\left(\alpha + \frac{\beta}{1+z}\right)}.$$

As $\alpha \to \infty$, the mass of the integral is increasingly concentrated in a neighborhood of the origin, say $z \in [0, \epsilon]$:

$$I_k(\alpha) \sim \int_0^\epsilon dz\, \frac{z^k}{1+z} e^{-z\left(\alpha + \frac{\beta}{1+z}\right)},\ \alpha \to \infty.$$

For any fixed $\epsilon > 0$, it results $\frac{\beta}{1+z} \in [\frac{\beta}{1+\epsilon}, \beta]$; therefore, as $\alpha \to \infty$ it also results

$$I_k(\alpha) \sim \int_0^\epsilon dz\, \frac{z^k}{1+z} e^{-z(\alpha+\beta)},\ \alpha \to \infty.$$

The above integral can be expressed in closed form for $k = 1$ and $k = 2$. The result follows by computing the limit of the ratio in (.36), which turns out not to depend on $\epsilon$.

## 4.D Proof of (4.17)

By explicitly computing the derivative in the definition of $\mathcal{S}_\infty$, we can rewrite it as follows:

$$\mathcal{S}_\infty = \beta \lim_{\alpha\to 0} \alpha I_\beta(\alpha), \tag{.37}$$

where

$$I_\beta(\alpha) := \int_0^\infty dz\, e^{-\alpha z} e^{-\beta \frac{z}{1+z}} \frac{z}{1+z}. \tag{.38}$$

The idea of the proof is to find upper and lower bounds on $\alpha I_\beta(\alpha)$, that match in the limit. To this end, observe that

$$e^{-\beta} \frac{z}{1+z} \leqslant e^{-\beta \frac{z}{1+z}} \frac{z}{1+z} \leqslant e^{-\beta}\left(1 + \frac{(\beta-1)^+}{z}\right). \tag{.39}$$

Hence, a lower bound is

$$\begin{aligned} \alpha I_\beta(\alpha) &\geqslant \alpha \int_0^\infty dz\, e^{-\alpha z} e^{-\beta} \frac{z}{1+z} \\ &= e^{-\beta}(1 - \alpha e^\alpha E_1(\alpha)) \\ &\to e^{-\beta}. \end{aligned} \tag{.40}$$

In order to compute the upper bound, split the domain of integration to avoid a singularity at $z = 0$ (cf. (.39)) as follows. For any $\epsilon > 0$, it results

$$\begin{aligned}\alpha I_\beta(\alpha) &\leqslant \alpha \int_0^\epsilon dz\, z + \alpha \int_\epsilon^\infty dz\, e^{-\alpha z} e^{-\beta}\left(1 + \frac{(\beta-1)^+}{z}\right) \\ &= \alpha\frac{\epsilon^2}{2} + e^{-\beta-\epsilon\alpha} + (\beta-1)^+\alpha E_1(\epsilon\alpha) \\ &\to e^{-\beta},\end{aligned} \tag{.41}$$

where for $z \in [0, \epsilon]$ we used a trivial upper bound for the integrand of $I_\beta(\alpha)$. The result follows from (.37), (.40) and (.41).

# 4.E Proof of Theorem 2

In this appendix, we compute the moments (4.23) and prove that convergence in probability to their mean holds.

The first remark is that matrix $\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*$ is diagonal:

$$\begin{aligned}\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^* &= \sum_{j=1}^K \boldsymbol{s}_j \boldsymbol{a}_j \boldsymbol{a}_j^* \boldsymbol{s}_j^* \\ &= \sum_{j=1}^K |a_j|^2 \boldsymbol{e}_{\pi_j}^K \boldsymbol{e}_{\pi_j}^{K*} \\ &= \sum_{i=1}^N \left(\sum_{j=1}^K \mathbb{1}_{\{\pi_j = i\}} |a_j|^2\right) \boldsymbol{e}_i^N \boldsymbol{e}_i^{N*},\end{aligned} \tag{.42}$$

where $\pi_k$ denotes the nonzero element of the signature $\boldsymbol{s}_k$ and $\boldsymbol{e}_i^n$ denotes the $i^{\text{th}}$ vector of the canonical basis of $\mathbb{R}^n$. In the last step, we move randomness from vectors to scalars, which will be shortly useful. Indeed, $m_L$ can be written as follows:

$$\begin{aligned}m_L &= \frac{1}{N}\operatorname{tr}(\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*)^L \\ &= \frac{1}{N}\sum_{i=1}^N ([\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*]_{ii})^L \\ &= \frac{1}{N}\sum_{i=1}^N \left(\sum_{j=1}^K \mathbb{1}_{\{\pi_j = i\}} |a_j|^2\right)^L.\end{aligned} \tag{.43}$$

Call the sum in parenthesis $S_i$:

$$S_i := \sum_{j=1}^K \mathbb{1}_{\{\pi_j = i\}} |a_j|^2. \tag{.44}$$

Hence, the expected value of $m_L$ is

$$\mathbb{E}[m_L] = \mathbb{E}[S_1^L] = M_{S_1}^{(L)}(0), \tag{.45}$$

where $M_{S_1}^{(L)}$ denotes the $L^{\text{th}}$ derivative of the MGF of $S_1$ [65]. It can be shown that $M_{S_i}(t) = (1 + \frac{t}{(1-t)N})^K$, hence

$$M_{S_i}^{(L)}(0) = \sum_{l=1}^{L} \binom{L-1}{l-1} \frac{L!}{l!} \frac{K!}{(K-l)!\, N^l}, \tag{.46}$$

which in the LSL becomes

$$\mathbb{E}[m_L] \to \sum_{l=1}^{L} \left\lfloor \begin{matrix} L \\ l \end{matrix} \right\rfloor \beta^l, \tag{.47}$$

where Lah numbers make their appearance $\left\lfloor \begin{matrix} L \\ l \end{matrix} \right\rfloor := \binom{L-1}{l-1} \frac{L!}{l!}$. Alternatively, $\mathbb{E}[m_L]$ can be expressed by using generalized Laguerre polynomials, which naturally appear in the Taylor expansion of the asymptotic MGF of $S_1$.

In order to prove convergence in probability, it is sufficient to show that $\mathrm{Var}[m_L] = \mathbb{E}[m_L^2] - (\mathbb{E}[m_L])^2 \to 0$. We have already found $\mathbb{E}[m_L]$. By using (.43) and (.44), $\mathbb{E}[m_L^2]$ can be expressed as follows:

$$\mathbb{E}[m_L^2] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} S_i^L\right)^2\right] \tag{.48}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} \mathbb{E}[S_i^{2L}] + \frac{1}{N^2}\sum_{i\neq j} \mathbb{E}[S_i^L S_j^L]. \tag{.49}$$

The first term is $O(1/N)$ because $\mathbb{E}[S_i^{2L}]$ is bounded in the LSL. The second term tends to $\mathbb{E}[S_1^L S_2^L]$. In order to show that this term becomes asymptotically equal to $\mathbb{E}[m_L]^2$, we can actually show more, namely $S_1$ and $S_2$ are asymptotically independent ($S_1 \perp S_2$).

To this end, interpret $S_i$ as the sum of a (random) number of weights $w_k := |a_k|^2$, namely $S_i = \sum_{k\in\mathscr{K}_i} w_k$ for $\mathscr{K}_i := \{k : \pi_k = i\} \subseteq [K]$. $\mathscr{K}_i$ is the subset of users who have chosen dimension $i$. Since the weights are i.i.d. random variables, the only source of dependence between $S_i$ and $S_j$ lies in the number of users who have chosen dimensions $i$ and $j$, respectively. These numbers are $K_i := |\mathscr{K}_i|$ and are not independent. Indeed, the vector $(K_1, K_2, \ldots, K_N)$ is distributed according to a Multinomial law with probabilities $(1/N, 1/N, \ldots, 1/N)$. In particular, the MGF of $(K_1, K_2)$ is

$$M_{K_1,K_2}(t_1, t_2) = \left(\frac{1}{N}(e^{t_1} + e^{t_2} + (N-2))\right)^K,$$

and tends in the LSL to

$$M_{K_1,K_2}(t_1, t_2) \to e^{\beta(e^{t_1}-1)} \cdot e^{\beta(e^{t_2}-1)},$$

where each term can be recognized as the MGF of a Poisson random variable with mean $\beta$. Since $K_1 \perp K_2$ asymptotically, also $S_1 \perp S_2$ from the independence of the weights.

## 4.F Verifying the Carleman Condition

Carleman's condition is $\sum_{k\geqslant 1} \bar{m}_{2k}^{-1/(2k)} = \infty$. We start off by upper bounding $\bar{m}_{2k}$ as follows:

$$\begin{aligned}\bar{m}_{2k} &= \sum_{l=1}^{2k} \left\lfloor \begin{matrix} 2k \\ l \end{matrix} \right\rfloor \beta^l \\ &\overset{\text{(a)}}{<} \sum_{l=1}^{2k} (2k-1)^{2k-l} \binom{2k}{l} \beta^l \\ &\overset{\text{(b)}}{\leqslant} (2k-1)^{2k-1} \sum_{l=1}^{2k} \binom{2k}{l} \beta^l \\ &\overset{\text{(c)}}{<} (2k-1)^{2k} (1+\beta)^{2k},\end{aligned} \tag{.50}$$

where (a) follows from the inequality $\left\lfloor \begin{matrix} 2k \\ l \end{matrix} \right\rfloor = (2k-1)!/(l-1)! = (2k-1)(2k-2)\dots(2k-(2k-l)) < (2k-1)^{2k-l}$, (b) derives from upper bounding $(2k-1)^{2k-l}$ with $(2k-1)^{2k-1}$, (c) is from the binomial formula by including in the sum the $l=0$ term. Therefore, $\bar{m}_{2k}^{1/(2k)} < (2k-1)(1+\beta)$, thus

$$\sum_{k\geqslant 1} \bar{m}_{2k}^{-1/(2k)} \geqslant \frac{1}{1+\beta} \sum_{k\geqslant 1} \frac{1}{2k-1} = \infty.$$

## 4.G Proof of (4.27)

It is convenient to represent $C^{\text{opt}}(\beta,\gamma)$ (in nats) as

$$C^{\text{opt}}(\beta,\gamma) = \sum_{k\geqslant 1} \frac{e^{-\beta}\beta^k}{k!} \int_0^\gamma k \exp(1/x)\, E_{1+k}(1/x) \frac{dx}{x}, \tag{.51}$$

which can be derived by differentiating in (5.6) under the integral sign with respect to $\gamma$ and integrating back after the integration with respect to $\lambda$. From the fundamental theorem of calculus, we have

$$\frac{\partial}{\partial\gamma} \int_0^\gamma \exp(1/x)\, E_{1+k}(1/x) \frac{dx}{x} = \exp(1/\gamma)\, E_{1+k}(1/\gamma) \frac{1}{\gamma},$$

which tends to 1 as $\gamma \to 0$, hence, by L'Hôpital's rule,

$$\lim_{\gamma\to 0} \frac{C^{\text{opt}}(\beta,\gamma)}{\gamma} = \lim_{\gamma\to 0} \frac{\partial C^{\text{opt}}(\beta,\gamma)}{\partial\gamma} = \sum_{k\geqslant 1} \frac{e^{-\beta}\beta^k}{k!} k = \beta. \tag{.52}$$

## 4.H Proof of (4.28)

The second derivative of $C^{\text{opt}}(\beta,\gamma)$ can be computed similarly to Appendix 4.G, which results in

$$
\begin{aligned}
-\lim_{\gamma\to 0}\frac{\partial^2}{\partial\gamma^2}C^{\mathrm{opt}}(\beta,\gamma) & \\
&= -\sum_{k\geqslant 1}\frac{e^{-\beta}\beta^k}{k!}\lim_{\gamma\to 0}\frac{\partial}{\partial\gamma}\exp(1/\gamma)\,E_{1+k}(1/\gamma)\frac{1}{\gamma} \\
&= \sum_{k\geqslant 1}\frac{e^{-\beta}\beta^k}{k!}k(1+k) = 2\beta+\beta^2. \quad (.53)
\end{aligned}
$$

The result follows by (.53) and (.52).

## 4.I Proof of (4.29)

Using (.51) and the fundamental theorem of calculus yields

$$
\lim_{\gamma\to\infty}\gamma\frac{\partial C^{\mathrm{opt}}_{\mathrm{lds}}}{\partial\gamma} = \sum_{k\geqslant 1}\frac{e^{-\beta}\beta^k}{k!}\lim_{\gamma\to\infty}k\exp(1/\gamma)\,E_{1+k}(1/\gamma) = \sum_{k\geqslant 1}\frac{e^{-\beta}\beta^k}{k!} = 1-e^{-\beta}. \quad (.54)
$$

# Chapter 5

# Theoretical analysis of Low-dense SC-NOMA over Rayleigh fading channels without CSI

In this chapter, upper and lower bounds on capacity of low-dense SC-NOMA are investigated in the context of 5G New Radio (5G-NR), under the worst hypothesis of no channel knowledge at the receiver, i.e. operation without channel state information. The impact of peculiar features of 5G-NR on capacity bounds of optimum receivers is investigated. The upper bound, defined as the system capacity with perfect CSI, and the lower bound corresponding to a pilot-based communication model, are found. These bounds indicate that the achievable rates of low-dense SC-NOMA without CSI are lower, as expected, than those with perfect CSI, although the above gap vanishes to negligible values, when some of the system parameters, in particular speed of variation of the channel against symbol duration, number of simultaneous users, and system load, i.e. number of users vs. number of resource elements, favorably combine.

## 5.1 Introduction

In recent real-world mobile communication channels, the rapid change of fading coefficients may severely affect estimation at the receiver. Receivers may have, in fact, no knowledge of channel state information (CSI), i.e. channel state information at the receiver (CSIR), except for the distribution of channel gains, a condition that is commonly referred to as *noncoherent setting* [69]. As a consequence, capacity, i.e. the fundamental criterion for predicting achievable rates, may be reduced compared to ideal conditions of perfect CSI detection, leading to a so-called noncoherent capacity [69, 70].

As a matter of fact, NOMA is a study item for 5G-New Radio (5G-NR), the standardization of which was launched by 3GPP in April 2016, with two phases aimed at defining 5G specifications before its commercialization in 2020. While other 5G

specifications are ongoing, the parameters of the radio framework were already specified in the very recent 3GPP specification TS 38.211 [71]. Understanding channel capacity for code-domain NOMA system in the absence of CSIR becomes, hence, essential. Closed-form expressions for noncoherent capacity are, however, difficult to obtain, even for the 'easy' memoryless channel [69], leading to a common practice to look for upper and lower capacity bounds, where the upper bound naturally corresponds to the capacity with perfect CSIR (coherent capacity) [72]. In order to find the lower bound, a popular method is to adopt a pilot-based communication model, for example, for the continuous fading model [23], and for the block-fading model [73].

This chapter investigates capacity bounds of low-dense SC-NOMA over Rayleigh fading channels without CSIR, based on the framework presented in [32]. The upper bound can be readily evaluated via the corresponding coherent capacity [32], whereas the lower bound can be derived via the capacity of a pilot-based communication scheme.

The chapter is organized as follows. Section 5.2 describes the channel and signal models. Capacity bounds are presented in Sec. 5.3. Results and discussion are reported in Sec. 5.4. Section 5.5 concludes the chapter.

## 5.2 System model

### 5.2.1 Channel model: Rayleigh block-fading assumptions

To reflect the nature of fading communication, either continuous or block-fading model can be selected to derive the capacity bounds with the duality property [73, 74]:

$$n_b = \frac{1}{2 f_m T_s}, \tag{5.1}$$

where $n_b$ represents the number of coherent symbols within a block in which the channel is considered stationary (block-fading model), $T_s$ is the symbol period, $f_m = v f_c / c$ is the maximum Doppler frequency, $f_c$ being the carrier frequency, $v$ is the velocity of interest, and $c$ is the speed of light. $n_b$ can be considered as the discretized version of the coherence time $T$ of a continuous-fading model. For 3GPP LTE and IEEE 802.16 WiMAX, the values of $n_b$ may range from unity to several hundreds [73], or even to thousands for 5G-NR as further detailed in Section 5.4.

### 5.2.2 Signal model

The model provided in [32] for low-dense NOMA was adopted, with the inclusion of a Rayleigh block-fading model. The channel model is rewritten as

$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{X} + \boldsymbol{N}, \tag{5.2}$$

where $\boldsymbol{Y} \in \mathbb{C}^{N \times n_b}$ is the received signal with $N$ referring to the number of resource elements (REs); $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K]^{\mathsf{T}} \in \mathbb{C}^{K \times n_b}$ with $\boldsymbol{x}_k$ being the row-vector including $n_b$ symbols transmitted by the $k$-th user; $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K] \in \mathbb{R}^{N \times K}$ is a random spreading matrix, column $k$ being the unit-norm spreading sequence of user $k$; $\boldsymbol{A} \in \mathbb{C}^{K \times K}$ is a

diagonal matrix of complex-valued fading coefficients $\{a_1, \ldots, a_K\}$; and $\boldsymbol{N} \in \mathbb{C}^{N \times n_b}$ denotes the noise with each column-vector $\boldsymbol{n} \in \mathbb{C}^N$ being described by a circularly-symmetric Gaussian random vector with zero mean and covariance matrix $\mathcal{N}_0 \boldsymbol{I}$. Compared to [32], vectors $\boldsymbol{y}$, $\boldsymbol{n}$, $\boldsymbol{x}$ are expanded with an additional dimension, since the reference unit is now a coherence block consisting of $n_b$ symbols rather than one symbol, whereas matrices $\boldsymbol{A}$ and $\boldsymbol{S}$ still hold in the new context.

The massive connectivity feature of 5G networks, that is, the large number of users that the system must provide with robust access simultaneously, is well reflected in a typical *asymptotic* analysis, known as system analysis in the asymptotic regime. As a matter of fact, the fundamental limits of NOMA and traditional DS-CDMA frameworks in the asymptotic regime were provided in [32] and [19] as further detailed in Section 5.3.

Noteworthy that the channel model in eq. (5.2) is equivalent to the general model $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}$ given in [75] for DS-CDMA, where $\boldsymbol{H} \in \mathbb{C}^{N \times K}$ is replaced by $\boldsymbol{S}\boldsymbol{A} \in \mathbb{C}^{N \times K}$. As stated by Tulino and Verdú in [75], the dimension roles of $N$ and $K$ of matrix $\boldsymbol{H}$ may assume different meanings for different systems; $N$ and $K$ are, for instance, the number of users vs. chips (REs) in DS-CDMA [19] and time-hopping CDMA [21], or the number of receiving ($n_R$) and transmitting ($n_T$) antennas in multiple-input multiple-output (MIMO). Therefore, results obtained for the different systems can be reused provided that a same signaling model holds, and the parameters are given the right meaning. In particular, the capacity lower bound for NOMA can be derived based on the analysis of lower bound for a MIMO system, provided by a pilot-based scheme [23].

## 5.3 Capacity bounds of low-dense NOMA

In this section, we find the optimum capacity bounds of low-dense NOMA in the noncoherent setting with the upper bound given by the capacity with perfect CSIR (coherent capacity) [32] and the lower bound based on the capacity of a pilot-based scheme originally proposed by Hassibi and Hochwald [23], and further refined by Rusek *et al.* [73].

### 5.3.1 Capacity upper bound

The ergodic capacity with perfect knowledge of the channel at the optimum receiver provided for the general model when $\boldsymbol{X}$ is a circularly-symmetric complex Gaussian with zero mean and covariance $\boldsymbol{Q}$, in bits/s/Hz, is provided in [76], and writes:

$$C = \mathbb{E}\left[\log_2 \det\left(\boldsymbol{I}_N + \boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^*\right)\right], \tag{5.3}$$

where $\boldsymbol{I}_N$ denotes the identity matrix.

For code-domain NOMA, eq. (5.3) becomes [32]:

$$C(\mathrm{SNR}) = \mathbb{E}\left[\log_2 \det\left(\boldsymbol{I}_N + \mathrm{SNR}\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{S}^*\right)\right], \tag{5.4}$$

whereas for single-user MIMO [75], the capacity is

$$C(\mathrm{SNR}) = \mathbb{E}\left[\log_2 \det\left(\boldsymbol{I}_N + \frac{\mathrm{SNR}}{K}\boldsymbol{H}\boldsymbol{H}^*\right)\right]. \tag{5.5}$$

The presence of coefficient $1/K$ in eq. (5.5) may be explained based on the fact that the transmitted power is divided equally among $K$ transmitting antennas, while in the NOMA model, each user transmits with full power (see eq. (5.4)).

The closed-form expression of coherent capacity for the particular case of low-dense NOMA in the asymptotic regime is as follows [32]:

$$C(\beta, \mathrm{SNR}) = \sum_{k \geqslant 1} \frac{e^{-\beta}\beta^k}{k!} \int_0^\infty \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} \log_2(1 + \mathrm{SNR}\lambda)\, d\lambda, \tag{5.6}$$

where $\beta$ is the system load, and the first term of right-hand side expression of eq. (5.6) is the probability density function (PDF) of a compound Poisson distribution with exponentially-distributed summands.

### 5.3.2 Capacity lower bound

In this section, capacity lower bound for low-dense NOMA is derived from the capacity of a pilot-based communication [23, 73]. First, a description of the pilot-based is provided, then, a follow-up discussion on optimizing the number of pilot symbols is presented, and two capacity lower bounds of low-dense NOMA corresponding to two different ways of allocating power to the pilots conclude the section.

#### Pilot-based channel model

Two phases typically characterize a pilot-based scheme: the pilot phase and the data transmission phase, where the latter is implemented after a minimum mean-square error (MMSE) channel estimation obtained during the pilot phase. Among the total $n_b$ symbols of a fading block, $n_p$ symbols, called pilot symbols, are allocated for learning the channel, and the remaining $(n_b - n_p)$ are dedicated for data transmission.

Three assumptions rule the pilot-based scheme proposed in this work. First, based on 5G properties as will be also indicated in Section 5.4, $n_b$ is much greater than both $\{K, N\}$, and it is, therefore, reasonable to assume $n_b > 2K$, i.e. the number of users is always lower than half the number of coherence symbols. Secondly, perfect channel estimation (via pilot phase) is assumed. Lastly, each transmission is assumed to be self-contained, that is, both pilot and data phases are referred to a specific fading block of $n_b$ symbols [73]; specific iteration of the procedure including training, estimation and data transmission is applied to each block.

*Pilot phase* For pilot symbols, one can write:

$$\boldsymbol{Y}_p = \boldsymbol{SAP} + \boldsymbol{N}_p, \tag{5.7}$$

where $\boldsymbol{P} \in \mathbb{C}^{K \times n_p}$ is the matrix of known pilot symbols replacing the transmitted signal $\boldsymbol{X}$ with the constraint $\boldsymbol{PP}^* = n_p\boldsymbol{I}$, $\boldsymbol{Y}_p$ and $\boldsymbol{N}_p$ are matrices of size $N \times n_p$.

*Data transmission phase* In this phase, a similar equation to eq. (5.2) can be applied with new dimensions of output and input being $N \times (n_b - n_p)$ and $K \times (n_b - n_p)$, respectively.

### Optimizing the number of pilot symbols

The number of pilot symbols $n_p$ directly affects the fundamental limits of a system in the noncoherent setting, and the issue of defining the optimal $n_p$ has been extensively investigated in the literature [23, 70, 73]. If $n_p$ is too small, the time dedicated to channel sounding may be insufficient to provide good estimates, while a too large $n_p$ value implies reduced data transmission rates [23].

An early study on the multi-antenna channel capacity by Marzetta [77], showed that to optimize the throughput of a pilot-based scheme, one should, in general, spend half of coherence time for training. The later work on multi-antenna capacity by Hassibi and Hochwald [23] provided a fine description of the optimal training interval as a function of the number of transmitting antennas $K$, receiving antennas $N$, fading coherence time $T$, and SNR. In fact, it was shown that in case of either low SNR or $T$ slightly larger than $K$, the pilot-based scheme is suboptimal; oppositely, when SNR is high or $T \gg K$, the lower bound capacity provided based on the pilot scheme approaches the coherent capacity. As detailed in Section 5.4, for NOMA with pilot-based channel estimation, capacity bounds can be analyzed under the impact of the number of symbols within a coherence block $n_b$, the number of users $K$, the number of REs $N$ (via the system load $\beta = K/N$) and SNR.

### The effect of pilot power allocation

Another condition also directly affecting capacity of pilot-based systems is pilot power allocation. In general, in order to have tight capacity bounds, a favorable condition is to avoid imposing constraints on power allocated to pilot symbols. Two capacity lower bounds are derived below, with no power constraint vs. with constraint.

*No constraint on pilot power allocation* If one can freely allocate power to pilot and data symbols (*pilot power-boosting* mode), the optimum number of pilot symbols is equal to the number of users, that is, $n_p = K$ [23, 70]. Intuitively, to learn the channel properly, the receiver should receive at least one pilot symbol from each user, i.e. $K$ pilot symbols in total. The lower bound of nonherent capacity of NOMA, given that the pilot power-boosting is allowed, can be directly inferred from [23, 73] with the aforementioned assumption of $n_b > 2K$, as follows:

$$C_{\mathrm{LB}} = \left(1 - \frac{K}{n_b}\right) C\left(\gamma_{\mathrm{eff}}\right), \tag{5.8}$$

where:

$$\gamma_{\mathrm{eff}} = \frac{n_b \mathrm{SNR}}{n_b - 2K} \left(\sqrt{\alpha} - \sqrt{\alpha - 1}\right)^2 \tag{5.9}$$

is the *effective* SNR replacing the 'regular' one of eq. (5.4, 5.6), and:

$$\alpha = \frac{n_b \mathrm{SNR} + K}{n_b \mathrm{SNR} \frac{n_b - 2K}{n_b - K}}. \tag{5.10}$$

From eq. (5.6), the closed-form expression for the lower bound capacity of NOMA is rewritten as follows:

$$C_{\mathrm{LB}}(\beta, \gamma_{\mathrm{eff}}) = \left(1 - \frac{K}{n_b}\right) \sum_{k \geqslant 1} \frac{e^{-\beta} \beta^k}{k!} \int_0^\infty \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} \log_2(1 + \gamma_{\mathrm{eff}} \lambda)\, d\lambda, \tag{5.11}$$

where $\gamma_{\mathrm{eff}}$ is taken from eq. (5.9).

*With constraint on pilot power allocation* If the pilot and data symbols are required to have the same power, then solving a convex optimization, depending on SNR, $n_b$, and $K$, leads to the following spectral efficiency, in bits/s/Hz [23]:

$$\max_{1 \leqslant n_p \leqslant n_b} \left(1 - \frac{n_p}{n_b}\right) C\left(\frac{\mathrm{SNR}^2 n_p / K}{1 + \mathrm{SNR}(1 + n_p/K)}\right). \tag{5.12}$$

## 5.4 Results and discussion

In this section, capacity bounds of low-dense NOMA, described in Sec. 5.3, are analyzed as a function of:

- number of coherence symbols $n_b$,
- number of users $K$,
- system load $\beta = K/N$.

The analysis of lower bound focuses on the case of no pilot power constraint (eq. (5.11)), the optimal number of pilot symbols, is thus equal to the number of users, i.e. $n_p = K$. An example with pilot power constraint (eq. (5.12)) is shown for reference in Sec. 5.4.3.

### 5.4.1 The impact of number of coherence symbols

The calibration of $n_b$ in the 5G context, is particularly relevant for New Radio (5G-NR). According to the duality property from eq. (5.1), factors that directly affect $n_b$ include the carrier frequency, the symbol period, and vehicular velocities. The new requirements and specifications of 5G has an impact on values of $n_b$ compared to the former generations such as 3GPP LTE [78], or IEEE 802.16 WiMAX [79], in particular:

- The carrier frequency $f_c$ varies within a wide range 1 – 100 GHz, with the deployment of macro sites at lower frequencies, and micro and pico sites at higher frequencies [80].
- Supported mobility speeds $v$ are up to 500 km/h, while the vehicular velocities of interest are about 120 km/h [80].

- According to 3GPP TS 38.211 [71], the most distinguished difference in frame structure of 5G-NR compared to LTE is the so-called numerology, i.e. the subcarrier spacing. In LTE, there is only one type of subcarrier spacing, that is 15 kHz, while multiple types of subcarrier spacing are supported in 5G-NR by scaling up in the power of 2, to the order of fifth numerology, that is, $2^5 \times 15 = 480$ kHz. Since the subcarrier spacing lies between 15 kHz and 480 kHz, its inverse is the symbol period $T_s$ ranging from $2\mu s$ to $66.7\mu s$.

Based on the relationship between $n_b$ and the above parameters, one obtains very high values of $n_b$ if the values of $f_c, v, T_s$ are low and vice versa. Taking into account typical values in the 5G-NR context, $n_b$ can vary from one to thousands, where its range in the earlier generations LTE or WiMAX was only from one to hundreds [73]. This increase is due to the fact that very large spacing of subcarrier in 5G-NR is also supported, leading to the improvement in the capacity of the pilot-based scheme (c.f. eq. (5.11)).

Figure 5.1 shows the capacity bounds of low-dense NOMA as a function of $E_b/N_0$, with upper bound defined by coherent capacity (solid line) and lower bounds obtained by a pilot-based scheme (via eq.(5.11)) with fixed $n_p = K = 10$ and different values $n_b = \{25, 50, 100, 500\}$, under the constraint $n_b > 2K$. The system load $\beta$ is kept fixed and equal to unity, i.e. $K = N$. Although the capacity bounds are considered in the asymptotics, where both $K$ and $N$ should be very large, the order of ten users or REs, however, was mentioned in [32, 67] is enough representative for a large-scale system. Note that, to show a fair comparison among systems, capacity bounds are shown below as a function of $E_b/N_0$ rather than of SNR via the relation $\frac{E_b}{N_0} = \beta.\text{SNR}/C(\text{SNR})$ [20].

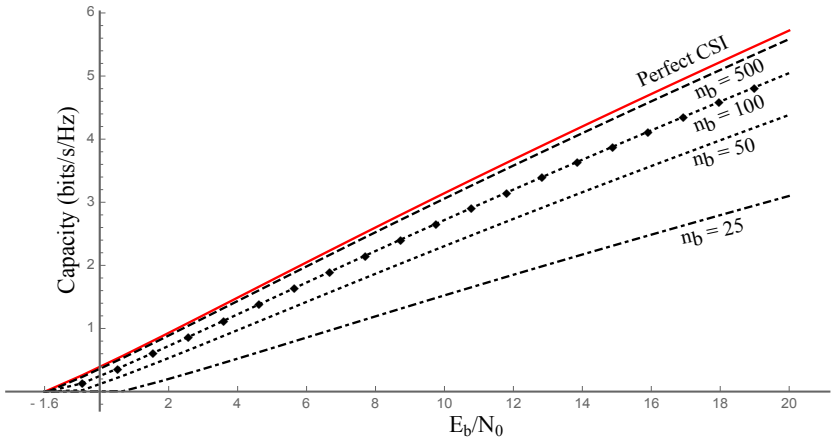



Figure 5.1. Capacity bounds (bits/s/Hz) of low-dense NOMA as a function of $E_b/N_0$ for fixed $\beta = 1$, with upper bound defined by coherent capacity (solid line) and lower bounds by a pilot-based scheme (eq. (5.11)) for different values $n_b$ ($n_b = 25, 50, 100, 500$). Note that the pilot power-boosting mode for the pilot scheme is adopted with fixed $n_p = K = 10$.

One may observe that given a same number of users, for e.g. $K = 10$, the scenario producing higher number of symbols in a coherence block benefits the tighter capacity bound to the perfect CSIR case. When $n_b$ is very high and $K$ is

fairly low ($n_b = 500, K = 10$), the gap between the upper and lower bounds becomes negligible. Remind that, such a high value of $n_b$ is, in fact, feasible in 5G-NR.

### 5.4.2 The impact of number of users

The impact of $K$ on the capacity lower bounds of low-dense NOMA, defined by the pilot-based scheme with pilot-boosting power (eq.(5.11)) is investigated with respect to the number of coherence symbols, for example, up to the favorable value $n_b = 500$. Figure 5.2 shows capacity lower bounds of low-dense NOMA as a function of $n_b$ with different values $K = \{25, 50, 100\}$, providing fixed $\beta = 1$ and $E_b/N_0 = 10$ [dB]. One may find that the higher the $K$ values, the lower the capacity of the pilot-based scheme, contrarily to MIMO, where the more the transmitting antennas $K$, the higher the system capacity; in NOMA since the more users, the more pilot symbols, less symbols are available for data transmission. The gap in capacity bounds between the lower $K = 10$ and higher cases $K = 100$ is thus substantial, as shown in Fig. 5.2.

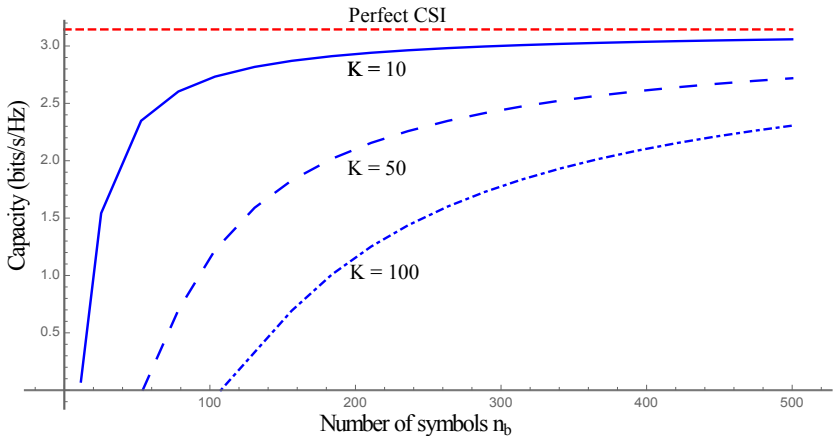



Figure 5.2. Capacity lower bounds (bits/s/Hz) of low-dense NOMA pilot-based scheme (eq.(5.11)) with fixed $n_b = 500$, $\beta = 1$ and $E_b/N_0 = 10$ [dB] as a function of $n_b$ with different values $K = \{10, 50, 100\}$.

### 5.4.3 The impact of system load

The system load $\beta = K/N$, which is an important factor in NOMA [32], leads to underloaded and overloaded systems when the number of users $K$ is lower (or $\beta < 1$) or higher than the number of REs (or $\beta > 1$). For optimum receivers, it was shown in [20, 21, 32] that the coherent capacity and the capacity lower bounds increase along with system load, irrespective of dense vs. low-dense spreading formats. Figure 5.3 displays the coherent capacity and its lower bounds given for the pilot-boosting power case as a function of $\beta$ for fixed $E_b/N_0 = 10$ [dB] in different $n_b$ and $K$ combinations, such as $\{n_b, K\} = \{25, 10\}, \{100, 10\}, \{500, 10\}, \{500, 100\}$. When $\beta$ increases, the capacity lower bounds with more coherence symbols (high $n_b$) and few users (low $K$) reach closer to the upper bound, as also observed in the $\beta = 1$

case. It is worthy to note that, the same lower capacity bounds can be attained given the same scaling of the combination $\{n_b, K\}$.

The capacity lower bound with the power constraint mode for the case $\{n_b = 25, K = 10\}$ is also shown in Fig. 5.3, and is, as expected, lower than that of without pilot power constraint, given the same combination $\{n_b, K\}$. From $\beta > 1$, the gap among the different cases expands substantially, particularly the lower bound derived for the case with pilot power constraint.

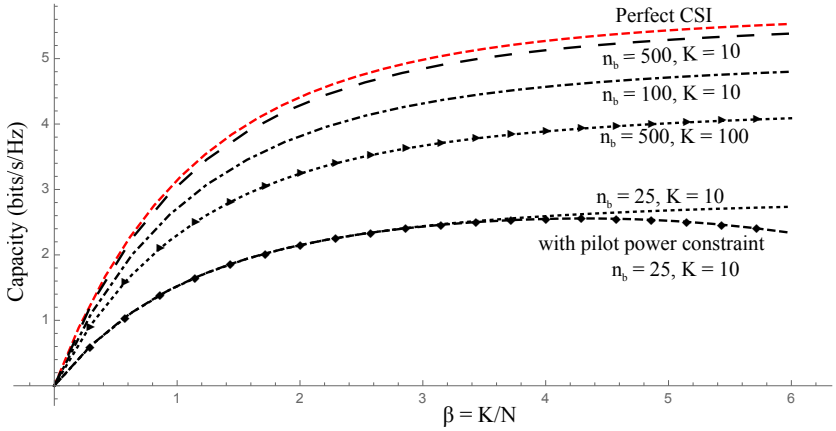



Figure 5.3. Capacity bounds of low-dense NOMA scheme with upper bound being the coherent capacity and lower bound obtained with a pilot-based scheme as a function of $\beta = K/N$ and fixed $E_b/N_0 = 10$ [dB].

## 5.5 Conclusion

Information-theoretical bounds of low-dense NOMA, under the hypothesis of Rayleigh block-fading channels without CSI, were derived. The capacity upper bound, defined as the capacity of low-dense NOMA with perfect CSI, was found based on the general framework proposed in [32]. The capacity lower bound was derived using a pilot-based communication scheme, as suggested in [23]. Upper and lower capacity bounds were described as a function of $E_b/N_0$, number of coherent symbols $n_b$, and system load $\beta$. The effect of the number of users $K$ was also investigated. Results indicate that, when the above factors are favourably combined, the gap between the upper and lower capacity bounds becomes negligible. In particular, when the number of symbols $n_b$ is high, while the number of users simultaneously joining the network is low, the lower capacity bound well approximates the capacity with perfect CSI, leading to the conclusion that the system is robust despite the absence of knowledge on the channel.

# Chapter 6

# Conclusions and Future works

Multiple access (MA) communication has been always present as a crucial research topic where many users may share the limited available resources under time-bandwidth constraints. Being acknowledged as an important enabler for 5G multiple access, NOMA with its diverse dialects recently attracted a huge attention in the wireless community from both industry and academia due to its massive connectivity, high bandwidth efficiency, ultra-low latency services, and heterogeneous data traffic.

The appearance of this thesis, which addressed the theoretical analysis of NOMA, and provided the latest research on different dialects of NOMA, is therefore very timely to provide a reference for the design of 5G multiple access. In this thesis, we focused on the classification of SC-NOMA from the well-known framework of CDMA, and on the fundamental limits of SC-NOMA under several practical issues including the impact of system features such as system load, sparsity, regularity, and the impact of channel, such as flat fading with and without CSI. Theoretical analysis of interest is all considered in large system limit, corresponding to massive communication, when both the number of users $K$ and the number of degrees of freedom $N$ grow unboundedly while its ratio $\beta = K/N$ is kept finite.

## 6.1 Conclusions

Below main conclusion are drawn from each chapter of the thesis.

In Chapter 1, an analytical framework, built on the traditional CDMA model, was introduced and was later proved to be flexible enough for representing several code-domain dialects. This chapter also addressed properties of a fundamental element of the model, that is, the representation matrix $\boldsymbol{S}$. Based on this framework, NOMA schemes were classified as single- vs. multi-carrier system due to the adopted waveform, and further divided into low-dense vs. dense group due to the property of the characterizing mapping matrix $\boldsymbol{S}$ of each NOMA dialect in Chapter 2.

In the remained chapters, single-carrier NOMA schemes were investigated in terms of spectral efficiency.

In Chapter 3, theoretical investigations of single-carrier NOMA were interpreted based on closed-form expressions existing for three distinctive cases in AWGN channels, that are, dense vs. *regular* low-dense and *irregular* low-dense NOMA,

corresponding to the model of DS-CDMA ($N_s = N$) [20] vs. LDS-CDMA ($N_s = 1$) [21] and *regular* sparse NOMA ($2 \leq N_s \in \mathbb{N}^+, \beta N_s \in \mathbb{N}^+$) [22]. Both optimum and linear receivers were taken into consideration. For any value of load, results shown that low-dense NOMA were more spectral-efficient than dense cases. In particular, spectral efficiency of the *regular* low-dense was always higher than the *irregular* low-dense and the dense NOMA with respect to optimum receivers. The price to pay is that one should allocate exactly $N_s$ REs per user and $\beta N_s \in \mathbb{N}$ users per resource, that is equivalently to have either central scheduling or users coordination. The theoretical behavior is distinctive for linear receivers: *regular* low-dense NOMA was proved to have highest achievable rates than all the other cases in the underloaded regime, while spectral efficiency of *irregular* low-dense overcome other NOMA cases in the overloaded systems (for e.g. when the system load $\beta \in [1.2,\ 5]$).

Chapter 4 and 5 focused on fundamental limits of *irregular* low-dense NOMA. In particular, Chapter 4 studied the impact of flat fading on theoretical analysis of *irregular* low-dense systems with linear and optimum receivers in the LSL. To this end, spectral efficiency was derived as a function of the load and signal-to-noise ratio. The framework used extended the model in [21], which was build on the seminal work [20], to the case with fading, along the lines of [19] in both the underloaded and overloaded regimes corresponding to $\beta < 1$ and $\beta > 1$, respectively. Given that the model in [21] originally defined for time hopping CDMA, it was shown that the model also fits the dual frequency domain in describing frequency-hopping CDMA. Since spreading sequences of *irregular* low-dense system are indeed the sparse counterparts of the dense ones, the reference model introduced in [21] for time hopping CDMA could also be adopted in the present context, in order to describe both dense and low-dense, where the spreading matrix $\mathbf{S} \in \mathbb{C}^{N \times K}$ of the model defines the sparsity of the system. Theoretical behavior of *irregular* low-dense system was compared against the dense given the the presence of flat fading was taken into consideration. One remark from [19] is that for dense case, the fading effect improves spectral efficiency in the overloaded condition. In this chapter, we analytically showed that, in the LSL, *irregular* low-dense NOMA with fading obtained higher gain in terms of spectral efficiency than the dense in the overloaded mode. The out-performance of low-dense case over the dense in terms of spectral efficiency goes stronger when $E_b/N_0$ increases.

In Chapter 5, a practical issue, that is no prior channel knowledge (CSI) at the receiver side, was tackled. Theoretical behavior of SC-NOMA, particularly *irregular* low-dense cases, was considered under this assumption. It is, in general, not easy to obtain the closed-form expressions of the so-called noncoherent capacity. Information-theoretical bounds of *irregular* low-dense NOMA under the hypothesis of Rayleigh block-fading channels without CSI were therefore derived, adopting the proposed framework (c.f. Chapter 4 [32]). Based on this setting, the capacity of low-dense NOMA with perfect CSI found in Chapter 4 turned out the capacity upper bound of interest. To find the capacity lower bound of *irregular* low-dense, we applied a pilot-based communication scheme, as suggested by [23]. The results illustrated upper and lower capacity bounds as a function of $E_b/N_0$, number of coherent symbols $n_b$, and system load $\beta$. The effect of the number of users $K$ was also studied. We showed that when the above factors are favorably combined, the lower capacity bound reaches the upper bound. In particular, when there is many symbols

consisting a block, i.e. $n_b$ is high, while the number of users simultaneously joining the network is low, the gap between the bounds is negligible. This implies that the system is still robust under some conditions despite the absence of knowledge on the channel.

In conclusion, results suggest that for overloaded systems, i.e. massive deployments, low-density spreading may be, at least theoretically, the right way forward, based on the increased capacity with respect to dense coding with linear receivers. This property holds despite fading. It is also shown that, with linear receivers, low-density NOMA achieves even higher performance in terms of spectral efficiency in the presence of flat fading, which is common in wireless communications.

## 6.2 Future works

While most of the current researches focus on power-domain NOMA, there are still many rooms and challenges to target for code-domain NOMA, particularly in 5G context. In this section, we draw several brief lines on code-domain NOMA that can be developed in the future.

- Future work may focus on refining the understanding of overloaded systems with a more general structure of the sparsity of spreading sequences, e.g., when $N_S > 1$. It is interesting to understand which value of $N_S$ represents the boundary between dense and low-dense systems, in terms of capacity, and more generally how the system behaves as a function of $N_S$.
- While it was shown that tight capacity bounds of low-dense SC-NOMA can be obtained in 5G NR scenario in Chapter 5, closed-form expressions of noncoherent capacity are expected to be derived for the verification with the bounds given by the pilot-based communication.
- Traditional receivers such as optimum and linear receivers were considered for all cases in theoretical analysis of NOMA while multi-user detection (MUD) techniques, for e.g. message passing algorithm (MPA)-based receivers, are suggested as low-complexity receivers for code-domain NOMA. Spectral efficiency of typical code-domain NOMA with respect to MPA-based detector is hence attracted to investigate.
- Theoretical study of this thesis focused on single-carrier NOMA. It is therefore appealing to have a fair comparison between SC-NOMA and MC-NOMA, particularly for low-dense signalling system, given that for dense format, capacity of MC-CDMA was shown lower than that of SC-CDMA [46].
- Theoretical study on hybrid scheme of power- and code-domain may open new direction for 5G multiple access design.

# Glossary

| | |
|---|---|
| 3GPP-LTE-A | 3rd generation partnership project long-term evolution advanced |
| CDMA | Code-division multiple access |
| CDM-NOMA | Code-domain non orthogonal multiple access |
| CR-NOMA | Cognitive radio non orthogonal multiple access |
| CSI | Channel state information |
| FDMA | Frequency division multiple access |
| FDS | Frequency domain spreading |
| IDMA | Interleave division multiple access |
| IDFT | Inverse discrete Fourrier transform |
| IGMA | Interleave-grid multiple access |
| IoT | Internet-of-thing |
| GOCA | Group orthogonal coded access |
| LCRS | Low code rate spreading |
| LDPC | Low-density parity-check |
| LDS-CDMA | Low density spreading code division multiple access |
| LDS-SVE | Low density spreading-signature vector extension |
| LSL | Large system limit |
| LSSA | Low code rate and signature based shared access |
| MA | Multiple Access |
| MIMO | Multiple-input multiple-output |
| MMSE | Minimum mean square error |
| MPA | Message passing algorithm |
| MUSA | Multi-user shared access |
| NCMA | Non-orthogonal coded multiple access |
| NOCA | Non-orthogonal coded access |
| NR | New Radio |

| | |
|---|---|
| OFDMA | Orthogonal frequency division multiple access |
| OMA | Orthogonal multiple access |
| PDMA | Pattern division multiple access |
| PDM-NOMA | Power-domain non orthogonal multiple access |
| SCMA | Sparse code multiple access |
| SUMF | Single user matched filter |
| TDMA | Time-division multiple access |
| TH-CDMA | Time-hopping code division multiple access |
| RDMA | Repetition division multiple access |
| RE | Resource element |
| RSMA | Resource spread multiple access |
| UL | Uplink |
| ZF | Zero forcing |

# List of Publications

**Journals**:

- **Mai T. P. Le**, Guido Carlo Ferrante, Tony Q.S. Quek, Maria-Gabriella Di Benedetto, "Fundamental Limits of Low-Density Spreading NOMA with Fading," IEEE Trans. on Wireless Commun., 2018, 17, (7), pp. 4648–4659.
- **Mai T. P. Le**, Guido Carlo Ferrante, Giuseppe Caso, Luca De Nardis, Maria-Gabriella Di Benedetto, "On Information-theoretic limits of Code-domain NOMA for 5G," IET Commun., 2018, Vol. 12 Iss. 15, pp. 1864–1871.
- S. Boldrini, L. De Nardis, G. Caso, **M. T. P. Le**, J. Fiorina, M.-G. Di Benedetto, "muMAB: A Multi-Armed Bandit Model for Wireless Network Selection," MDPI Algorithms, 2018, vol. 11 (2), 13; doi: 10.3390/a11020013.
- G. Caso, **M. T. P. Le**, L. De Nardis, M.-G. Di Benedetto, "Performance Comparison of WiFi and UWB Fingerprinting Indoor Positioning Systems," MDPI Technologies, 2018, vol. 6 (1), 14; doi:10.3390/technologies6010014.
- G. Caso, L. De Nardis, **M. T. P. Le**, F. Maschietti, J. Fiorina, M.-G. Di Benedetto, "Performance Evaluation of Non-prefiltering vs. Time Reversal prefiltering in distributed and uncoordinated IR-UWB Ad-Hoc networks," Springer Mobile Networks and Applications (MONET), 2017, vol. 22 (5), pp. 796-805; doi: 10.1007/s11036-017-0829-6.

**Book Chapters**:

- G. Caso, **M. T. P. Le**, L. De Nardis, M.-G. Di Benedetto, "Non-Cooperative and Cooperative Spectrum Sensing for 5G Cognitive Networks," in W. Zhang (eds.), Handbook of Cognitive Radio, Springer International Publishing, 2017.

**Conference Proceedings**:

- **M. T. P. Le**, G. Caso, L. De Nardis, A. Mohammadpour, G. Tucciarone, M.-G. Di Benedetto, "Capacity bounds of Low-Dense NOMA over Rayleigh fading channels without CSI," in IEEE Proc. Int. Conf. on Telecom. (ICT), pp. 428–432, St. Malo, France, 26–28 June 2018.

# Bibliography


[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] L. Dai, B. Wang, Y. Yuan, S. Han, C. l. I, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE J. Sel. Areas Commun.*, vol. 53, no. 9, pp. 74–81, Sept. 2015.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[4] "5G Waveform & Multiple Access Techniques," Qualcomm Technologies, Inc., Nov. 2015.

[5] Q. Wang, L. L. Zhang, R.and Yang, and L. Hanzo, "Non-orthogonal multiple access: A unified perspective," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 10–16, 2018.

[6] G. Mazzini, "Power division multiple access," in *Proc. IEEE Int. Conf. on Universal Personal Commun. (ICUPC'98)*, vol. 1, Oct. 1998, pp. 543–546.

[7] T. Cover and J. Thomas, *Elements of information theory*, 2nd ed. New York, USA: John Wiley & Sons Inc., 2012.

[8] J. Choi, "On HARQ-IR for Downlink NOMA Systems," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3576–3584, Aug. 2016.

[9] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition Coding Strategies: Design and Experimental Evaluation," *IEEE Trans. Commun.*, vol. 11, no. 7, pp. 2628–2639, July 2012.

[10] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[11] R. Hoshyar, F. Wathan, and R. Tafazolli, "Novel Low-Density Signature for Synchronous CDMA Systems over AWGN channel," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1616–1626, Apr. 2008.

[12] R. Razavi, R. Hoshyar, M. Imran, and Y. Wang, "Information theoretic analysis of LDS scheme," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 798–800, Aug. 2011.

[13] J. Van De Beek and B. Popovic, "Multiple Access with Low-Density Signatures," in *Proc. IEEE Glob. Telecommun. Conf. (GLOBECOM)*, Honolulu, HI, 2009, pp. 1–6.

[14] R. Razavi, A. Mohammed, M. Imran, R. Hoshyar, and D. Chen, "On receiver design for uplink low density signature OFDM (LDS-OFDM)," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3499–3508, Nov. 2012.

[15] M. Zhao, S. Zhou, W. Zhou, and J. Zhu, "An Improved Uplink Sparse Coded Multiple Access," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 176–179, Jan. 2017.

[16] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern Division Multiple Access- A Novel Nonorthogonal Multiple Access for Fifth-Generation Radio Networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3185–3196, Apr. 2017.

[17] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, Nanjing, 2016, pp. 1–5.

[18] S.-Y. Lien, S.-L. Shieh, Y. Huang, B. Su, Y.-L. Hsu, and H.-Y. Wei, "5G New radio: Waveform, Frame Structure, Multiple Access, and Initial Access," *IEEE Trans. Commun.*, vol. 55, no. 6, pp. 64–71, June 2017.

[19] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[20] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[21] G. C. Ferrante and M.-G. Di Benedetto, "Spectral efficiency of random time-hopping CDMA," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6643–6662, Dec. 2015.

[22] B. M. Zaidel, O. Shental, and S. Shamai, "Sparse NOMA: A Closed-Form Characterization," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Vail, CO, USA, June 2018, pp. 1106–1110.

[23] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[24] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5g networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[25] J. A. F. Ross and D. P. Taylor, "Vector assignment scheme for $m + n$ users in $n$-dimensional global additive channel," *Electronics Letters*, vol. 28, no. 17, pp. 1634–1636, 1992.

[26] ——, "Multiuser signaling in the symbol-synchronous AWGN channel," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1174–1178, 1995.

[27] 3rd Generation Partnership Project (3GPP) Portal, 2017, [Online]. Available: https://portal.3gpp.org/, accessed on May 09, 2018.

[28] J. Zeng, B. Li, X. Su, L. Rong, and R. Xing, "Pattern division multiple access (PDMA) for cellular future radio access," in *Proc. IEEE Int. Conf. on Wireless Commun. and Signal Process. (WCSP)*, Nanjing, China, 2015, pp. 1–5.

[29] 3rd Generation Partnership Project (3GPP), *Non-orthogonal multiple access candidate for NR*, TSG RAN WG1 #85, Nanjing, China, May 2016, proposed by Samsung.

[30] K. Li, X. Wang, and L. Ping, "Analysis and optimization of interleave-division multiple-access communication systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, May 2007.

[31] L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, 2006.

[32] M. T. P. Le, G. C. Ferrante, T. Q. S. Quek, and M.-G. Di Benedetto, "Fundamental Limits of Low-Density Spreading NOMA with Fading," *IEEE Trans. Wireless Commun.*, 2018.

[33] X. Dai, S. Chen, S. Sun, S. Kang, Y. Wang, Z. Shen, and J. Xu, "Successive interference cancelation amenable multiple access (SAMA) for future wireless communications," in *Proc. IEEE Int. Conf. Commun. Sys. (ICCS)*, Macau, 2014, pp. 222–226.

[34] J. Choi, "Low density spreading for multicarrier systems," in *2004 IEEE Eighth International Symposium on Spread Spectrum Techniques and Applications*, 2004, pp. 575–578.

[35] L. Wen, R. Razavi, M. Imran, and P. Xiao, "Design of joint sparse graph for ofdm (jsg-ofdm) system," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 1823–1836, Apr. 2015.

[36] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE Annu. Int. Symp. Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, London, UK, 2013, pp. 332–336.

[37] 3rd Generation Partnership Project (3GPP), *Low code rate and signature based multiple access scheme for New Radio*, TSG RAN1 #85, Nanjing, China, May 2016.

[38] ——, *Considerations on DL/UL multiple access for NR*, TSG RAN WG1 #84bis, Busan, Korea, Apr. 2016.

[39] ——, *Non-orthogonal multiple access for New Radio*, TSG RAN WG1 #85, Nanjing, China, May 2016, proposed by Nokia, Alcatel-Lucent Shanghai Bell.

[40] ——, *New uplink non-orthogonal multiple access schemes for NR*, TSG RAN WG1 #86, Gothenburg, Sweden, Aug. 2016.

[41] J. Zhang, X. Wang, X. Yang, and H. Zhou, "Low density spreading signature vector extension (lds-sve) for uplink multiple access," in *Proc. IEEE Veh. Technol. Conf. (VTC-Spring)*, Toronto, Canada, 2017, pp. 1–5.

[42] S. D. Sosnin, G. Xiong, D. Chatterjee, and Y. Kwak, "Non-Orthogonal Multiple Access with Low Code Rate Spreading and Short Sequence Based Spreading," 2017, pp. 1–5.

[43] 3rd Generation Partnership Project (3GPP), *Initial views and evaluation results on non-orthogonal multiple access for NR uplink*, TSG RAN WG1 #84bis, Busan, Korea, Apr. 2016.

[44] ——, *Candidate NR multiple access schemes*, TSG RAN WG1 #84bis, Busan, Korea, Apr. 2016.

[45] X. Dai, Z. Zhang, S. Chen, S. Sun, and B. Bai, "Pattern Division Multiple Access (PDMA): A New Multiple Access Technology for 5G," *IEEE Wireless Commun. Mag.*, vol. 25, no. 2, pp. 54–60, Apr. 2018.

[46] A. M. Tulino, L. Li, and S. Verdú, "Spectral efficiency of multicarrier CDMA," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 479–505, Feb. 2005.

[47] A. Montanari and D. N. C. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove tanaka's formula)," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Punta del Este, 2006, pp. 160–164.

[48] M. Yoshida and T. Tanaka, "Analysis of sparsely-spread CDMA via statistical mechanics," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2006, pp. 2378–2382.

[49] J. Raymond and D. Saad, "Sparsely spread CDMA- a statistical mechanics-based analysis," *J. Phys. A: Math. Theor.*, vol. 40, no. 41, pp. 12 315–12 333, 2007.

[50] D. Guo and C.-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 421–431, Apr. 2008.

[51] Q. Xiong, C. Qian, B. Yu, and C. Sun, "Advanced NoMA Scheme for 5G Cellular Network: Interleave-Grid Multiple Access," in *Proc. IEEE Glob. Telecommun. Conf. (GLOBECOM)*, Singapore, 2017, pp. 1–5.

[52] 3rd Generation Partnership Project (3GPP), *Multiple access schemes for new radio interface*, TSG RAN WG1 #84bis, Busan, South Korea, Apr. 2016.

[53] M. Al-Imari, M. A. Imran, and P. Xiao, "Radio Resource Allocation for Multicarrier Low-Density-Spreading Multiple Access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2382–2393, Mar. 2017.

[54] 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, 2015.

[55] E. Paolini, C. Stefanovic, G. Liva, and P. Popovski, "Coded random access: applying codes on graphs to design random access protocols," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 144–150, 2015.

[56] O. Shental, B. M. Zaidel, and S. S. Shitz, "Low-density code-domain NOMA: Better be regular," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2017, pp. 2628–2632.

[57] B. M. Zaidel. Personal communications, 2018.

[58] D. N. C. Tse and S. V. Hanly, "Linear Multiuser Receivers: Effective Interference, Effective Bandwidth and User Capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.

[59] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 171–188, Jan. 2000.

[60] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, 2002.

[61] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[62] ——, "Capacity region of Gaussian CDMA channels: The symbol-synchronous case," in *Proc. Allerton Conf. on Commun., Control, and Computing (Allerton)*, Oct. 1986, pp. 1025–1034.

[63] V. L. Girko, *Theory of Random Determinants*. Dordrecht, Netherlands: Kluwer Academic, 1990.

[64] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. New York, USA: Springer, 2010, vol. 20.

[65] W. Feller, *An introduction to probability theory and its applications*. New York, USA: John Wiley & Sons Inc.„ 1968, vol. 2.

[66] L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions (Revised and Enlarged Edition)*. Dordrecht, Holland and Boston, USA: D. Reidel Publishing Co., 1974.

[67] D. Guo and S. Verdú, "Randomly Spread CDMA: Asymptotics via Statistical Physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, June 2005.

[68] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, USA: Dover Publications, 1972.

[69] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai, "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 367–395, Jan. 2010.

[70] L. Zheng and D. N. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[71] 3rd Generation Partnership Project (3GPP), *TS 38.211. NR; Physical channels and modulation*, 2017, [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/38211.htm.

[72] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[73] F. Rusek, A. Lozano, and N. Jindal, "Mutual information of IID complex gaussian signals on block rayleigh-faded channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 331–340, Jan. 2012.

[74] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.

[75] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Commun. and Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[76] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *Transactions on Emerging Telecommunications Technologies*, vol. 10, no. 6, pp. 585–595, 1999.

[77] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. Allerton Conf. on Commun., Control, and Computing (Allerton)*, vol. 37, 1999, pp. 958–966.

[78] *UTRA-UTRAN Long Term Evolution (LTE)*. 3rd Generation Partnership Project (3GPP), Nov. 2004.

[79] J. G. Andrews, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall PTR, Nov. 2007.

[80] K. Haneda and *et al.*, "Indoor 5G 3GPP-like channel models for office and shopping mall environments," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 694–699.