This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

Sparse Code Multiple Access : Potentials and Challenges

Manel Rebhi, Kais Hassan, Kosai Raoof and Pascal Chargé

Abstract—**The massive connectivity is among other unprecedented requirements which are expected to be satisfied in order to follow the perpetual increase of connected devices in the era of internet of things. In contrast to the family of conventional orthogonal multiple access schemes, the key distinguishing feature of non-orthogonal multiple access (NOMA) is its capacity to support the massive connectivity. Sparse code multiple access (SCMA) is one of the powerful schemes of code-domain NOMA (CD-NOMA) and is among the promising candidates of multiple access techniques to be employed in future generations of wireless communication systems thanks to the sparsity pattern of its codebooks. This technique has been actively investigated in recent years. In this paper, we provide a comprehensive survey of the state-of-the-art of SCMA. First, we will pinpoint SCMA place in the NOMA landscape including power-domain NOMA and CD-NOMA with the aim of justifying why SCMA is prominent. Then, its system architecture is highlighted and its basic principles are presented, afterwards a review of exiting codebook designs and available SCMA detectors is provided, before showing how resources are expected to be assigned, and how SCMA can be combined with other existing and emerging technologies. Finally, we present a range of future research trends and challenging open issues that should be addressed to optimize SCMA performance.**

Index Terms—**SCMA, NOMA, code-domain, codebook design, multi-dimensional constellations, message-passing algorithms, SCMA detector, IoT.**

Nomenclature

1G	First-Generation
2G	Second-Generation
3G	Third-Generation
3GPP	Third-Generation Partnership Project
4G	Fourth-Generation
5G	Fifth-Generation
ADC	Analog-to-Digital Converter
APSK	Amplitude and Phase Shift Keying
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BICM	Bit-Interleaved Coded Modulation
BP	Belief Propagation
BS	Base Station
CB	Codebook
CD-NOMA	Code-Domain NOMA
CDMA	Code-Division Multiple Access
CS	Compressive Sensing
CSI	Channel State Information
DMPA	Discretized MPA
DNN	Deep Neural Network
EML	Extended Maximum Likelihood
EPA	Expectation Propagation Algorithm
EXIT	Extrinsic Information Transfer
FBER	Frozen Bit Error Rate
FDMA	Frequency-Division Multiple Access
FFT	Fast Fourier Transform
FN	Function Node
FUO	Fixed User Order
GA-MPA	Gaussian-Approximated MPA
GenA-MPA	Generalized Approximate MPA
IDMA	Interleave Division Multiple Access
IGMA	Interleave Grid Multiple Access
IoT	Internet of Things
ISD	Improved SD
LDPC	Low-Density Parity Check
LDS-CDMA	Low-Density Spreading Code Multiple Access
LDS-OFDM	Low-Density Signature Orthogonal Frequency-Division Multiplexing
LLR	Log Likelihood Rate
LSD	List Sphere Decoding
LSTM	Long Short-Term Memory
LTE	Long-Term Evolution
MIMO	Multiple-Input Multiple-Output
MMSE-PIC	Minimum Mean Square Error with Parallel Interference Cancellation
mMTC	massive Machine Type Communication
mmWave	millimeter Wave
ML	Maximum Likelihood
MPA	Message Passing Algorithm
MSD	Modified Sphere Decoding
MS-MPA	Multiple Scheduling MPA
MSTS	Modified Single Tree Search
MUSA	Multi-User Shared Access
NOCA	Non-Orthogonal Coded Access
NOMA	Non-Orthogonal Multiple Access
NR	New Radio
OA	Opportunistic Assignment
OFDMA	Orthogonal Frequency-Division Multiple Access
OMA	Orthogonal Multiple Access
PA-MPA	Partially Active Message Passing Algorithm
PD-NOMA	Power-Domain NOMA
PDF	Probability Density Function
PDMA	Pattern-Division Multiple Access
PF	Proportional Fairness

M. Rebhi, K. Hassan and K. Raoof are with Le Mans University, France, 72085 Le Mans, LAUM, France, 72085 Le Mans, (e-mail: manel.rabhi.tn@ieee.org, kais.hassan@univ-lemans.fr, kosai.raoof@univ-lemans.fr).

P. Chargé is with Ecole polytechnique Université de Nantes, IETR, Nantes, France (e-mail: pascal.charge@univ-nantes.fr).

PM-MPA	Partial Marginalization Message Passing Algorithm
PSMA	Power-Domain Sparse Multiple Access
QAM	Quadrature Amplitude Modulation
Q-MPA	Quantum-Assisted MPA
QoE	Quality of Experience
QSD-MPA	Quantum-Assisted Sphere Decoder based MPA
RE	Resource Element
RIS	Reconfigurable Intelligent Surface
RPMA	Random Phase Multiple Access
RSMA	Resource Spread Multiple Access
S-NOMA	Signature-based NOMA
SAMA	Successive Interference Cancellation Amenable Multiple Access
SC-EPA	Sparse-Channel based EPA
SCMA	Sparse Code Multiple Access
SD	Sphere Decoding
SDMA	Spatial-Division Multiple Access
SIC	Successive Interference Cancellation
SISO	Soft-Input Soft-Output
SNR	Signal-to-Noise Ratio
S-SCL	Soft-Successive Cancellation List
SS-MPA	Single Scheduling MPA
TDMA	Time-Division Multiple Access
VN	Variable Node

I. Introduction and Background

The next decades will encounter many emerging applications which will require to massively connect new devices. For instance, the aim of the internet of things (IoT) is to massively connect various types of physical and virtual objects into a dynamic global network infrastructure and to enable applications such as connected cars, industrial IoT, smart cities, connected healthcare, etc [1]. The major targets of the fifth-generation (5G) mobile communication systems and beyond are the massive connectivity, the ultra-high data rates, the ultra-low latency, as well as flexible air interface design to support several user requirements [2]. The 5G is expected to increase the connectivity density by 10 (at least $10^6/\text{km}^2$) [3]. In 2021, the number of IoT devices and connected mobile users will exceed 50 billion of active hosts [4]. It will be very difficult to satisfy this huge demand on data traffic and massive connectivity without proposing new enhanced communications techniques especially enhanced multiple access schemes.

One key of the evolution of wireless communication systems over their different generations was to propose new multiple access techniques including frequency-division multiple access (FDMA) for first-generation (1G), time-division multiple access (TDMA) for second-generation (2G), code-division multiple access (CDMA) for third-generation (3G), and orthogonal frequency-division multiple access (OFDMA) for fourth-generation (4G) [5], [6]. All these techniques are orthogonal, that is the wireless resources are orthogonally shared among multiple users in different domains as time, frequency and code. The orthogonality of the allocated resource elements (REs) reduces the multiple access interference and had enabled low-complexity and efficient receivers. The aforementioned orthogonal multiple access (OMA) techniques are not sufficient to reply to the increasing connectivity demand of the future generations since the number of supported users is limited by the number of available orthogonal REs. In order to enable the massive access, non-orthogonal multiple access (NOMA) mechanisms were introduced [7]. The idea is to further exploit the REs by employing a non-orthogonal access, for instance, multiple users can share the same narrow frequency band over the same time slot which allows to serve more users at the expense of increasing the receiver's complexity. The NOMA techniques are divided into two main categories according to the type of allocated resources, namely: power-domain NOMA (PD-NOMA) and code-domain NOMA (CD-NOMA). However, other existing techniques can also assign resources belonging to multiple domains to multiple users in a non-orthogonal manner. Hence, the objectives of 5G and beyond motivated, and will continue to motivate, the research evolution of NOMA techniques despite the fact that the third-generation partnership project (3GPP) decided not to include them in 5G.

Recently, trying to incorporate PD-NOMA in 5G networks have gained attention of researchers around the globe [8]–[10]. The main idea of PD-NOMA is to exploit a new dimension, that is the power. On the other hand, CD-NOMA was inspired by the classic CDMA systems in which multiple users share the same RE by adopting a specific signature per user. The key difference between CD-NOMA techniques and CDMA ones is that the spreading sequences of the former are restricted to non-orthogonal low cross-correlation sparse sequences.

Sparse code multiple access (SCMA) is among the CD-NOMA techniques that are considered by the new radio (NR) study in the 3GPP long-term evolution (LTE)-advanced, since it was among the most reliable multiple access candidates for 5G [11], [12]. The SCMA encoder directly maps the incoming bits of several users/streams to complex multi-dimensional codewords selected from a specific predefined sparse codebook set. The shaping gain of the multi-dimensional constellation of SCMA leads to a better spectral efficiency. Due to the attractivity of SCMA and the active research around it, we think that it is important to provide a comprehensive survey that aims to explain its birth, to analyze its various approaches, to highlight its main advantages and to summarize the future research directions which can improve its performance and facilitate its integration in future mobile systems.

Contributions of the survey

Several papers were proposed to review existing NOMA mechanisms [6], [8], [13], [14], [14]–[24]. Some of these surveys focus only on PD-NOMA techniques [6], [8], [16]–[18], [24]. The authors in [21] provides a comprehensive survey of the interplay between PD-NOMA and some emerging technologies such as massive multiple-input multiple-output (MIMO), millimeter wave (mmWave), energy harvesting, wireless caching, and so on. In [23], PD-NOMA applications to wireless networks, including cellular networks, device-to-device communications and wireless sensor networks, was

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

- **Introduction and Background**
- **Non-Orthogonal Multiple Access**
  - Power-Domain NOMA
  - Code-Domain NOMA
  - NOMA Standardization in 3GPP
  - Pros & Cons of SCMA
- **SCMA System Architecture**
  - Basic Principles of SCMA
  - Irregular SCMA
- **SCMA Codebook Design**
  - Mother Constellation Design
  - Transformation Operators Design
- **SCMA Detector Design**
  - MPA and its Variations
  - Approximated Algorithms
  - Sphere Decoding
  - Detectors for Channel-Coded SCMA
  - Deep Learning based Detectors
  - MIMO-SCMA Detectors
- **Some Other SCMA Aspects**
  - Resource Allocation
  - SCMA Implementations
  - Interplay of SCMA
- **Open Directions of Research**

Figure 1. The outline of this survey.

reviewed. Also, recent advances in NOMA from a grant-free access perspective were comprehensively discussed in [22]. On the other hand, the authors in [13], [15], [19], [20] tried to cover the PD-NOMA and the CD-NOMA at the same time. Hence, each paper have presented briefly some aspects of SCMA. An overview of the existing multi-dimensional constellations for uplink SCMA systems was proposed in [14], this is the only survey that is dedicated to SCMA, however only codebook design is covered. Table I summarizes the contributions of existing surveys in a chronological order with a special attention on how SCMA is involved in each one among them.

The aim of this survey is to propose an inclusive overview of the existing SCMA mechanisms which can be employed as a part of future wireless systems and IoT networks. To the best of our knowledge, a specific review that discusses comprehensively SCMA does not exist. However, we think that such a survey is needed and will be useful.

Scope and Organization

The outline of this paper can be defined as :

- First, we briefly review different types of NOMA schemes in Section II which will help to better understand the importance of all of them and to locate SCMA in the large NOMA landscape by paying a close attention to its advantages and drawbacks.
- Afterwards, the SCMA system model is presented in Section III. We will highlight its architecture and its basic principles before covering the different approaches to design SCMA transceiver. This includes the SCMA codebook design in Section IV as well as the different existing detectors in Section V.
- Then, in the following section, we will turn our focus to other aspects such as SCMA resources allocation, interplay between SCMA and other emerging technologies, and practical experiments and prototyping of SCMA.
- Finally, discussion and open directions of research are summarized in Section VII. The paper then, is concluded in Section VIII.

The structure of our survey is illustrated in Figure 1.

Notation

Lightface letters denote scalars. Boldface lower- and upper-case letters denote column vectors and matrices, respectively. $[\mathbf{A}]_{n,m}$ and $\mathbf{x}_i$ stand for the entries of matrix $\mathbf{A}$ and vector $\mathbf{x}$, respectively. $\mathbf{A}^T$ is the transpose of $\mathbf{A}$, and $\mathbf{A}^*$ is the optimum solution of an optimization problem. Moreover, $\text{diag}(\mathbf{x})$ is a diagonal matrix whose diagonal elements are the entries of $\mathbf{x}$. $\mathbf{A}_j$ and $\mathbf{x}_j$ represents, respectively, a matrix and a vector of user j and $\{\boldsymbol{A}_j\}$ denotes a set of matrices of user j. ϕ is the SCMA system. $\mathbf{x} \backsim \mathcal{CN}(0, N_0\mathbf{I}_K)$ denotes that the $K \times 1$ vector $\mathbf{x}$ is a zero-mean circularly complex Gaussian variable with variance N_0 where $\mathbf{I}_K$ is the identity matrix of size K. $\mathbb{P}(.)$ is the probability of an event and $\mathbb{E}(.)$ is the expected value of a random variable. $\mathcal{O}(*)$ describes the complexity order.

II. Non-Orthogonal Multiple Access

The beyond 5G wireless networks face various challenges in order to support large scale heterogeneous traffic and users, therefore new modulation and multiple access schemes are being developed to meet the growing demand [30]. Based on how the resources are shared among multiple users, two types of multiple access can be distinguished: OMA and NOMA. The non-orthogonality means that users can exploit simultaneously resources belonging to two different domains at least, for instance it could be sharing the same time slot and the same subcarrier for uplink (or downlink) communications while still be able to decode the message of each user despite the inter-user interference. NOMA, which has been recently proposed for the 3GPP LTE-advanced [6], constitutes hence a promising technology for addressing the aforementioned challenges in beyond 5G networks by accommodating several users within the same RE. By doing so, significant bandwidth efficiency enhancement can be attained over conventional OMA techniques.

NOMA schemes are mainly divided into two categories according to how users are multiplexed at available REs, namely PD-NOMA and CD-NOMA [15]. Figure 2 gives a classification of existing multiple access schemes. Apart

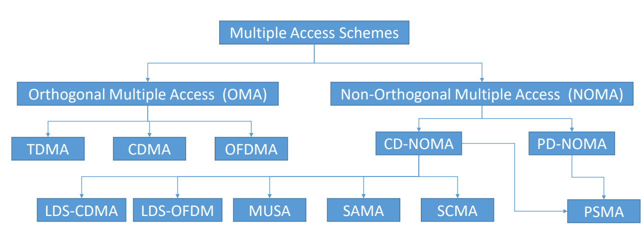

Figure 2. An overview of exiting multiple access schemes

Table I
REVIEW OF EXISTING SURVEYS ON NOMA WITH A FOCUS ON HOW EACH SURVEY COVERS DIFFERENT SCMA ASPECTS.
$\surd \rightarrow$ DISCUSSED IN DETAILS, $\approx\rightarrow$ BRIEFLY DISCUSSED, $\times$ $\rightarrow$ NOT DISCUSSED

Reference	Year	Contributions	SCMA			
			Basic Principles	Codebook Design	Detector Design	Other Aspects
Our proposed SCMA Survey: SCMA techniques, SCMA encoder and detector designs and some SCMA aspects			$\surd$	$\surd$	$\surd$	$\surd$
Shirvanimoghaddam *et al.* [25]	2016	Review of NOMA techniques for massive cellular IoT solutions, random access and resource allocation strategies for M2M communications.	$\times$	$\times$	$\times$	$\times$
Wei *et al.* [7]	2016	Survey of downlink PD-NOMA: Design, performance and key features.	$\times$	$\times$	$\times$	$\times$
Ding *et al.* [13]	2017	A survey on PD-NOMA, cognitive radio NOMA, multi-carrier NOMA, cooperative NOMA, mmWave NOMA, practical implementation.	$\approx$	$\times$	$\times$	$\times$
Islam *et al.* [8]	2017	Review of PD-NOMA: Basics concepts, capacity analysis, power allocation strategies, user fairness, user-pairing, and interplay with cooperative MIMO, beamforming and space–time coding.	$\times$	$\times$	$\times$	$\times$
Liu *et al.* [6]	2017	Survey of PD-NOMA: Basic principels, muti-user detection, interference cancelation, NOMA with multiple antennas techniques (beamforming, MIMO), cooperative NOMA communications, resource management, multi-carrier NOMA and standardization.	$\approx$	$\times$	$\times$	$\times$
Benjebbour *et al.* [26]	2017	Overview of PD-NOMA: Concept, benefits, issues, MIMO-NOMA, performance evaluation with an experimental trial.	$\times$	$\times$	$\times$	$\surd$
Shan *et al.* [27]	2017	Uplink NOMA review: a short summary of existing PD-NOMA and CD-NOMA schemes.	$\surd$	$\times$	$\times$	$\times$
Dai *et al.* [15]	2018	A survey of NOMA: Basic principles and advantages, PD-NOMA solutions, CD-NOMA solutions, other NOMA schemes, user grouping and resource allocation, performance evaluations and prototypes.	$\surd$	$\times$	$\times$	$\approx$
Ding *et al.* [18]	2018	PD-NOMA : General principles, application to other advanced communication technologies (wireless caching, MIMO techniques,mmwave communications, and cooperative networks).	$\approx$	$\times$	$\times$	$\times$
Cai *et al.* [20]	2018	Overview of advances on modulation and multiple access techniques, PD-NOMA, CD-NOMA, performance comparison.	$\surd$	$\times$	$\times$	$\approx$
Vameghestahbanati *et al.* [14]	2018	Overview of existing SCMA codebooks: origin, multidimensional constellation design, key performance indicators and performance comparison.	$\surd$	$\surd$	$\approx$	$\times$
Ye *et al.* [19]	2018	Typical NOMA technologies, grant-free NOMA (motivation, typical schemes and detection techniques), implementation issues.	$\surd$	$\times$	$\times$	$\approx$
Aldababsa *et al.* [16]	2018	A tutorial about basic concepts of PD-NOMA technique and its usage in MIMO and cooperative scenarios along with some practical implementation aspects and performance analyses.	$\times$	$\times$	$\times$	$\times$
Basharat *et al.* [17]	2018	A taxonomy of NOMA schemes, different objectives, associated constraints and solution approaches for optimization problems of NOMA, decoding methods and key performance indicators.	$\approx$	$\times$	$\times$	$\times$
Ma *et al.* [28]	2019	A chapter on SCMA: non-comprehensive tutorial which deals with basic description, performance evaluation, codebook design and SCMA applications.	$\surd$	$\approx$	$\approx$	$\approx$
Anwar *et al.* [23]	2019	A survey about applications of PD-NOMA to cellular networks, device-to-device communications, and wireless sensor networks: network model, performance evaluation for each application.	$\surd$	$\times$	$\times$	$\surd$
Vaezi *et al.* [21]	2019	Review of NOMA combinations with other technologies with an emphasis on how these technologies interplay and benefit from each other: MIMO, mmWave communications, cooperative communications, cognitive radio, physical layer security, energy harvesting, etc.	$\times$	$\times$	$\times$	$\times$
Shahab *et al.* [22]	2019	First comprehensive review of the recent advances in NOMA from a grant-free connectivity perspective: Grant-free NOMA schemes with their potential and related practical challenges for mMTC/IoT and possible future directions.	$\approx$	$\times$	$\times$	$\approx$
Makki *et al.* [29]	2020	Review of 3GPP study items on NOMA and their related decisions, link-level performance evaluation, UE pairing, receiver design and NOMA-HARQ.	$\approx$	$\times$	$\times$	$\times$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

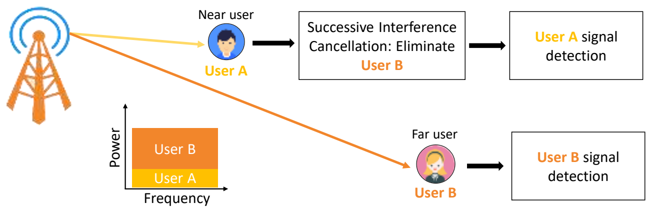

Figure 3. Illustration of downlink PD-NOMA principles. User A (near user) is with better channel conditions, while user B (far user) is with poorer channel conditions.

from these two categories, other existing alternative NOMA schemes was also proposed in the state-of-the-art. In the following subsections, we will briefly present the key NOMA technologies by addressing the different domains of non-orthogonality such as power, code, and other ones. For more details on these techniques, their performance, and available measurements and experimental results, reader can refer to survey papers mentioned in Table I. Afterwards, a rough comparison of the two NOMA families, in terms of their spectral efficiency, system performance, receiver complexity, will be introduced. This will help the reader to possess a global summary of multiple access techniques and consequently to understand why SCMA gained a lot of attention.

A. Power-Domain NOMA

The basic idea of PD-NOMA, so-called NOMA in several works as well, is to allow supporting more multiple access by employing a new dimension : the signal's power [8]. PD-NOMA is based on combining two key techniques namely, superposition coding at the transmitter, and mainly successive interference cancellation (SIC) at the receiver's side to segregate user information. These two components are briefly presented in the following paragraphs.

Superposition Coding: The fundamental concept of superposition coding is to guarantee higher detection reliability by superimposing several users at the same RE with a power allocation among users based on the near-far propriety. The idea is to allocate more power when the user is far from the base station (BS), i.e. when the normalized channel gain is smaller. Multiplexing users in PD-NOMA consists of superimposing the constellation diagram of different users such that each user's information is modulated and summed among users with appropriate power allocations so that the resulting signal still form higher-order constellation.

Successive Interference Cancellation: At the receiver, SIC can be invoked by the users to cancel interference. The optimal detection order is from the strongest user to the weakest one such that each user can detect its message without substantial interference imposed by the stronger users whose signals are already detected and removed from the superposed signal. In this way, we first decode the message of the user with the most allocated power where the messages from other ones are considered as noise, and thus the decoded message is least contaminated. This mechanism is depicted in Figure 3.

Theoretically, PD-NOMA is optimal in terms of achieving the capacity boundary [15]. However, despite the fact that the SIC receiver exhibits a low complexity when compared to other CD-NOMA receivers, it is inherently sensitive to error propagation, i.e. once an error occurs in decoding a user's signal, the rest of users' data will be corrupted and decoded erroneously. This implies keeping the number of users reasonably low to reduce the effect of error propagation such that a marginal impact on the PD-NOMA performance is granted [31]. On the other hand, the error propagation problem may lead to a considerable retransmission overhead and consequently a high packet transmission delay [32]. Another limit to the PD-NOMA schemes is the high cost of dynamic power allocation and users pairing especially when the number of users increases which results in a huge feedback overhead and requires some complex optimization methods [9], [33]. Different implementation aspects of SIC-based receiver for PD-NOMA systems were reviewed in [29], the authors had also proposed some advanced solutions for automatic-repeat-request based data transmission [34] and simplified user pairing [35]. As regards hardware aspects, the analog-to-digital converter (ADC), employed to quantify the signal at SIC receiver, suffers from the dynamic range problem [36]. That is, more the strength of the received signal is different from one user to another, more dynamic range of ADC is needed for quantization, and more realizing an efficient solution is difficult.

B. Code-Domain NOMA

The concept of CD-NOMA, which had been inspired by the classic CDMA systems, allows multiple users to share the same RE (time-frequency), but each user adopts unique user specific spreading sequences which presents its signature. However, separating these users at the receiver requires sophisticated techniques, such as SIC and message passing algorithm (MPA), at the cost of an increased complexity. According to the different designs of code, we can distinct several CD-NOMA schemes in addition to SCMA, the main existing ones are presented in what follows.

Low-Density Spreading Code-Division Multiple Access (LDS-CDMA): The most basic scheme of CD-NOMA is the LDS-CDMA [37], [38] which extends the classical CDMA. In CDMA systems, each user is associated to one among non-sparse quasi-orthogonal spreading sequences. Each spreading sequence is composed of a number of time slots so-called *chips*. The symbol of all users are spread over different chips and superimposed before transmission. At the receiver, a simple low-complexity correlation detector is sufficient to cancel the inter-sequence interference thanks to the orthogonality of sequences, but this comes at the expense of the number of connected users. LDS-CDMA aims to increase the number of users by spreading their information over a small number of chips rather than all of them which consequently limits the interference on each chip. Hence, low-density and consequently non-orthogonal sequences are required. The MPA detector is needed at the receiver even when the additive white Gaussian noise (AWGN) assumption is considered. Hence, LDS-CDMA has the drawback of high multi-user detection complexity.

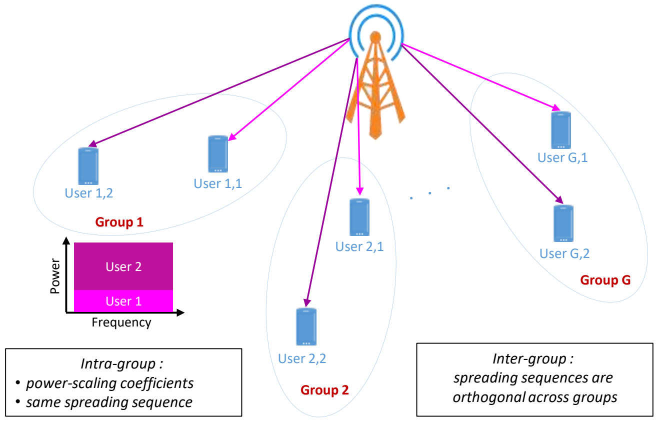

Figure 4. Basic principles of MUSA.

Low-Density Signature Orthogonal Frequency-Division Multiplexing (LDS-OFDM): LDS-OFDM can be interpreted as a transformed version of LDS-CDMA where each user's symbol is spread across a carefully selected number of subcarriers and superimposed in the frequency domain which makes LDS-OFDM more adapted for strong frequency-selective channels [39].

Multi-User Shared Access (MUSA): In a downlink scenario, MUSA users are divided into several groups. In each group, the signal of each user is multiplied by its specific power scaling coefficient. Next, the signals of all the users of a given group are added and then the superimposed signal is spreaded through a specific sequence [40] as presented in Figure 4. The group-specific sequences are designed to be orthogonal such that inter-group interference is easily eliminated at the receiver by using a low-complexity correlation. Finally, a SIC detector can be employed to solve the intra-group interference problem.

Successive Interference Cancellation Amenable Multiple Access (SAMA): Unlike MUSA, SAMA divides users into several groups based on a joint design of the system signature matrix and iterative multi-user interference cancellation technique such that a diversity gain is obtained [41]. The manner in which users are organised into groups must accelerate the convergence of the iterative detector at the receiver by first eliminating the most reliable users (the ones whose data is spreaded on more subcarriers) at each iteration which will consequently facilitate decoding the less reliable users. The authors in [41] considered that the above-described design possesses a convergence-amenable property. An iterative MPA can be efficiently exploited at the receiver.

Hybrid multiple access: Few existing works tried to combine SCMA with other multiple access techniques [42]–[44]. Kim *et al.* [42] proposed a hybrid system based on PD-NOMA and SCMA where near and far property of PD-NOMA is used to serve multiple users simultaneously. The data of a group of nearest users is superimposed using SCMA codebooks, while a number of far users which is equal to the number of REs is supported such that each user is associated to one subcarrier. Later, in order to improve the spectral efficiency, power domain sparse code multiple access (PSMA) extends SCMA by allowing the same SCMA codebook to be used simultaneously by multiple users, then the signals of these users are non-orthogonally multiplexed in the power-domain before transmitting the superimposed signal on different subcarriers [43], that is the overloading is performed both in the power and code domain. Recently, another coexistence between PD-NOMA and SCMA was introduced in [44] based on flexible joint non-uniform power allocation and multi-dimensional code sparsity, this scheme allows to increase the number of supported users. For these schemes, an MPA based SIC detector is employed to decode the transmitted signals.

C. Other-Domains NOMA

Signature-based NOMA (S-NOMA) schemes are closely-related to CD-NOMA ones since each user is distinguished by a specific signature which can be based on codebook design, delay pattern, interleaving pattern, spreading and scrambling sequence [45]. The signature represents how data bits are spread over available REs in non-orthogonal manner. Motivated by the error propagation problem caused by SIC in PD-NOMA, the pattern-division multiple access (PDMA) [46], [47] uses a specific sparse pattern to map transmitted data onto small part of a group of resources, i.e. the signature of PDMA is its codebook structure. The PDMA pattern is a binary vector whose length is the number of resources in the group, the binary value of each one among its entries determines if the user is mapped to the corresponding resource. The patterns are selected to maximize the diversity order while minimizing the overlapping among users [45]. Belief propagation (BP) based iterative detection and decoding is adopted at the receiver in an uplink scenario, while BP or SIC method is applied in downlink scenario. Another multiple access approach is to employ delay patterns such that the transmitted frames do not arrive exactly at the same time. For example, random phase multiple access (RPMA) [48] guarantees the inter-group separation by assigning the same Gold code to all the users in the same group, however, the superimposed signals inside a given group can be segregated since they are transmitted asynchronously after a random delay.

Other alternative NOMA candidates are based on interleaving patterns where different bit-level interleavers are employed for user separation with low multiple user interference. This category includes interleave grid multiple access (IGMA) [27] and interleave division multiple access (IDMA) [49]. IGMA achieves coding and diversity gains by randomly distributing the transmitted bits based on a combination of bit-level interleaver and sparse symbol-level grid mapping patterns which adds another dimension to user multiplexing process. A special case of IGMA is IDMA where different users can be distinguished by bit-level interleavers combined with low-rate error-correction codes. Another S-NOMA approach is based on the scrambling sequences which allows high degrees of freedom for user separation using specific scrambling sequence for each user. One scheme in this category is resource spread multiple access (RSMA) [50], [51] which employs a combination of very low-rate error-correction codes and long enough scrambling sequences with good correlation properties to well separate signals. We can also mention the non-orthogonal

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

coded access (NOCA) scheme [52] which spread data symbols before transmission in frequency and/or time domain via low correlation sequences which are defined by LTE, hence it will not be very complicated to include NOCA in the 3GPP LTE standard. Minimum mean-square error (MMSE) based NOCA receiver outperforms existing OFDMA schemes in link level performance.

Some other domains were also investigated to propose a range of alternative NOMA schemes such as spatial-division multiple access (SDMA) [53]–[55]. This advanced multiple antenna technique aims to separate users by assigning them to different spatial signatures in a MIMO system through creating multiple spatial sub-channels in order to increase the degree of freedom. However, SDMA performance is limited through frequency-selective channels.

The NOMA techniques are extensively reviewed in [8], [15], [21] and [24].

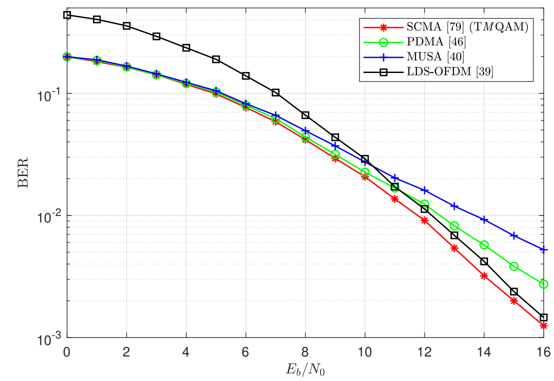

Figure 5. Performance comparison of LDS-CDMA, MUSA, PDMA and SCMA, in terms of BER, through Rayleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6.

D. NOMA Standardization progress in 3GPP

In this subsection our aim is to review all technical outcome and specification set proposed by the 3GPP, especially by its technical specification group *Radio Access Network* (TSG RAN), to investigate NOMA for long-term evolution (LTE) (under Release 12, Release 13, Release 14) and for 5G (under Release 15 and Release 16). Since 2013, 3GPP considered the non-orthogonality pattern in different applications. For example, a non-orthogonal extension of the network-assisted intercell interference mitigation in LTE Release 12 was introduced [56]. Then, since their 87$^{\text{th}}$ meeting, NOMA has been selected as a study item in LTE Release 13, in which, PD-NOMA was termed as multi-user superposition transmission (MUST) [57]. The study concluded that PD-NOMA can increase system capacity and improve user experience since multiple users' transmissions can be superimposed on each other leading to enhanced spectral efficiency. By the end of 2014, NOMA was also selected as a work item in LTE Release 14. Furthermore, many CD-NOMA schemes were evaluated for massive machine type communications (mMTC) scenario, and the results show significant benefits of CD-NOMA in terms of uplink sum throughput and overloading capability, as well as system capacity enhancement in terms of supported packet arrival rate at given system outage. In April 2016, a discussion on different multiple access techniques for both downlink and uplink was proposed in [58] where OMA schemes (TDMA, FDMA and SDMA) and NOMA ones (MUSA, LDS-CDMA, SCMA and RSMA) were compared in order to provide candidates for the three different types of services for 5G networks whose requirements vary according to supported functionalities: mMTC (improved link budget, low device complexity, low energy consumption, high density device deployment), enhanced mobile broadband small data scenarios (eMBB) (low latency, higher spectral efficiency/throughput) and ultra-reliable low latency communication (URLLC) (low packet error rate, low latency). Then, by the end of 2016, the activity on NOMA as work item, was put on hold to give way for more basic 5G functionalities. Many NOMA schemes were evaluated in LTE Release 14 and the decision was to continue with NOMA as study item for 5G Release 15 at least for mMTC. Later, due to the strong interest in both academia and industry [59], NOMA study was revived in Release 15 for both URLLC and eMBB scenarios [60]. That is NOMA was studied under four main aspects: I) transmission side NOMA schemes, II) receiver algorithms, III) NOMA related procedures and IV) preliminary performance evaluations.

Furthermore in Release 16, 3GPP TSG RAN, during their 85$^{\text{th}}$ meeting in May 2016, considered (a) 3 categories of CD-NOMA based on how data is spread over allocated resources: symbol-level spreading such as MUSA, bit-level scrambling/interleaving such as IDMA, and joint modulation and spreading such as SCMA, and (b) 3 major typical multi-users receivers: MMSE hard interference cancellation, elementary signal estimator with soft-input soft-output (SISO) decoder, and expectation propagation algorithm (EPA) with SISO decoder.

Several proposed NOMA schemes in 3GPP studies were categorized and compared, in a qualitative manner, in [61]. Recently, the authors in [29] had also highlighted the final conclusions presented in NOMA study-items of 3GPP Release 15. No clear gain from NOMA over Release 15 and Release 16 mechanisms was observed in all studied scenarios. Mainly, in a large number of cases, link-level results from many companies showed no gain, and the system-level simulations do not provide conclusive performance enhancement. Currently, NOMA techniques are not deployed in the 5G, and are not mentioned explicitly in the planning of 3GPP Release 17 and Release 18. However, NOMA is expected to be reconsidered in very high density scenarios of beyond 5G generations [29].

E. Pros & Cons of SCMA

Among all the existing non-orthogonal technologies, SCMA scheme is shown to achieve a better link level performance [62]. Unlike LDS-OFDM, bit to constellation mapping and low-density spreading are combined together, i.e. bits directly mapped to different sparse codewords in SCMA.

Comparison study between PD-NOMA and SCMA in [63] showed that SCMA outperforms PD-NOMA when comparable resources allocation strategies for heterogeneous cellular

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

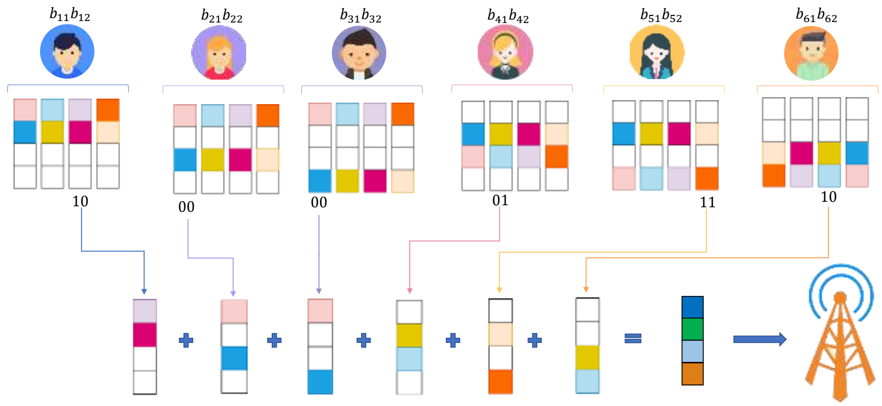

Figure 6. Representation of the encoder of a regular SCMA system.

networks are employed. This gain in performance is at the expense of more complexity where SIC method is used to separate signals in the PD-NOMA case while the more sophisticated MPA one is used for SCMA systems, the later costs more in terms of operations and materials [23]. That is why a lot of research work was conducted to reduce the complexity of SCMA detectors as it will be shown in Section IV, however additional future contributions are expected to further enhance this aspect. SCMA employs a multi-dimensional codebook which leads to a constellation shaping gain and consequently to better spectral efficiency when compared to other CD-NOMA schemes as LDS-CDMA and LDS-OFDM [64]. The authors in [65] studied the problem of random access in the mMTC where a large number of IoT devices hope to get access. It was shown that SCMA outperforms the current LTE random access procedure and that of PD-NOMA in terms of access success probability, average access delay and throughput [65], [66]. In [67], the performance of SCMA wireless networks was examined via a tractable analytical framework using tools from stochastic geometry. This work studied the impact of SCMA codebook design on the network performance especially the optimal transmitter density.

Figure 5 illustrates a performance comparison among some NOMA techniques, namely LDS-OFDM, MUSA, PDMA and SCMA, in terms of bit error rate (BER), through Rayleigh fading channel. Simulation results confirm that SCMA scheme provides better performance when compared to others.

III. SCMA System Architecture

In this section, we want to highlight the basic principles and proposed schemes of SCMA in order to explain how it can provide multiple access.

A. Basic principles of SCMA

We consider a synchronous SCMA system with a BS and J separate users so-called layers and K OFDM subcarriers, so-called REs. A SCMA transmitter encodes $\log_2(M_j)$ data bits of user j and maps them into a K-dimensional codeword, $\mathbf{x}_j$, which is selected from a distinct codebook $\boldsymbol{C}_j$ of size M_j. Each codebook presents the signature of the corresponding user. The codebooks are built based on multi-dimensional constellation, $\mathcal{C} = \{\boldsymbol{C}_j, 1 \leq j \leq J\}$, whose shaping gain enables it to outperform the traditional spread code based schemes [68]. As to the codewords of SCMA, they are sparse, i.e. only $N_j \ll K$ of their entries are non-zero and the rest are zeros, N_j is called *codebook sparsity degree*. The sparsity key of SCMA that all codewords corresponding to the j^{th} SCMA layer have a unique location of non-zero entries at the same $(K - N_j)$ positions. A regular SCMA system is defined by $N_j = N, 1 \leq j \leq J$ and $M_j = M, 1 \leq j \leq J$, an example is presented in Figure 6 where all users employ a codebook of size 4 and their signals are spread over two REs, i.e. $M_j = 4, N_j = 2, 1 \leq j \leq 6$. The maximum degree of user superposition on a given RE is denoted d_f, and the overloading factor, λ, is defined by the ratio of number of users to number of REs, $\frac{J}{K}$. Regrading the system in Figure 6, $d_f = 3$ and $\lambda = 150\%$. The codewords of all users, $\mathbf{x}_j, 1 \leq j \leq J$, are superimposed and exchanged over the K REs.

The K-dimensional uplink received vector is given by,

$$\mathbf{y} = \sum_{j=1}^{J} \mathbf{H}_j \mathbf{x}_j + \mathbf{n}, \tag{1}$$

where $\mathbf{y} = (y_1, \cdots, y_K)^T$, $\mathbf{x}_j = (x_{j,1}, \cdots, x_{j,K})^T$, $\mathbf{H}_j = \text{diag}(\mathbf{h}_j)$ and $\mathbf{h}_j = (h_{j,1}, \cdots, h_{j,K})^T$ is the $K \times 1$ channel gain vector of user j. The $K \times 1$ vector $\mathbf{n}$ corresponds to the additive zero-mean white circularly complex Gaussian noise with variance N_0; i.e. $\mathbf{n} \backsim \mathcal{CN}(0, N_0\mathbf{I}_K)$, where $\mathbf{I}_K$ is the identity matrix of size K. However, Equation (1) is transformed to express the K-dimensional downlink received vector of user u, $\mathbf{y}_u$, as follows,

$$\mathbf{y}_u = \mathbf{H}_u \sum_{j=1}^{J} \mathbf{x}_j + \mathbf{n}. \tag{2}$$

The constellation function, associated with each user j generates a constellation set with M_j alphabets of length N_j. Then, the mapping matrix V_j maps the N_j-dimensional constellation points to SCMA codewords to form the codebook $\boldsymbol{C}_j$. A SCMA system can be described by a $K \times J$ factor graph matrix $\mathbf{F} = (\mathbf{f}_1, \cdots, \mathbf{f}_J)$ whose columns define the positions of non-zero element of each user. Hence, the matrix $\mathbf{F}$ is related to the codeword $\mathbf{x}_j$, in equation (1) and equation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

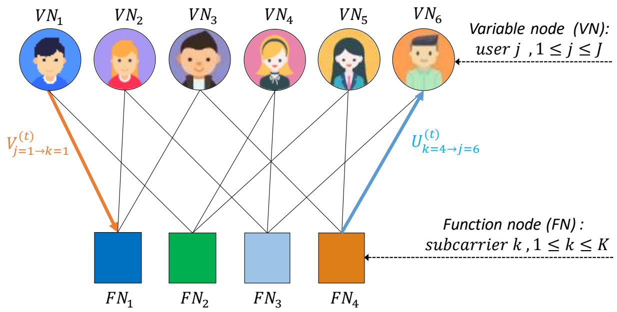

Figure 7. MPA factor graph corresponding to the SCMA system presented in Figure 6.

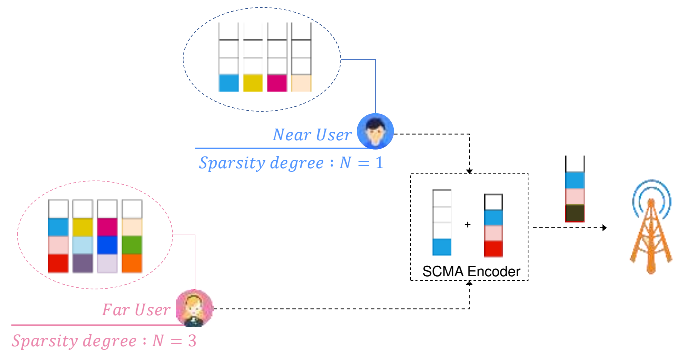

Figure 8. Irregular uplink SCMA system with near and far users using different sparsity degrees, $N = 1$ and $N = 3$ respectively, based on their channel states [69].

(2), by the fact that the structure of $\mathbf{F}$ defines where zeros are located in the codebook from which the codeword $\mathbf{x}_j$ is selected. Another way to represent the system is to employ $\mathbf{V} = \{\mathbf{V}_j, 1 \le j \le J\}$ where $\mathbf{V}_j$ is the mapping matrix of user j, it is worth noting that $\mathbf{f}_\mathrm{j} = \mathbf{V}_j\mathbf{V}_j^T$. For instance, the system in Figure 6 is represented by,

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$\Longrightarrow \text{For instance, } \mathbf{V}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

A factor graph is depicted in Figure 7 where every circle represents a user (so-called variable node) and every block represents a subcarrier (so-called function node).

B. Irregular SCMA

Most existing SCMA propositions are based on regular structure where users are treated equally that is why these propositions can not be adapted to users' business needs or conditions in the same scenario, this is not ideal. So new irregular SCMA systems are needed to manage different requirements such as capacity, connectivity, and data rate, and to also exploit the knowledge of channel state information (CSI) [69]. Some existing works tried to solve this problem, for instance an irregular SCMA codebook design was proposed in [70] to assign different codebooks with various dimensions according to different users' requirements. Simulations show that using an irregular codebook design does improve the system performance in terms of BER, however the authors did not take into consideration the impact of the correlation among users. Another irregular SCMA design was proposed in [71], the idea was to employ different rotated angles to design different codebooks for several user's needs, nevertheless the proposed codebooks are still far from being optimal. The authors in [72] studied the resource allocation for different users in the same system by proposing a flexible resource scheduling scheme. Other contribution in [73] proposed an energy-saving algorithm for a joint codebook design and assignment, and power allocation for both uplink and downlink SCMA scenarios. In [69], the total performance of the system is adjusted according to users' channel states by allocating more REs to the users with the worst channel conditions. The users are divided into different groups according to their CSI and the near-far strategy is applied as shown in Figure 8. This adaptive SCMA scheme outperforms the regular one when different channel states are assumed which is more realistic.

In order to enhance SCMA mapping system, researchers in [74] proposed a more flexible uplink SCMA scheme by directly mapping a variable number of coded symbols from each user onto subcarriers. This will improve the overloading factor and system performance at the expense of more complex signaling control.

IV. SCMA Codebook Design

Here, our aim is to highlight existing techniques to conceive SCMA codebooks. The SCMA codebook design is a joint optimization problem of the optimum user-to-RE mapping matrix $\mathbf{V}^*$ and the optimum multi-dimensional constellation $\mathcal{C}^*$, and can be defined as,

$$\mathbf{V}^*, \mathcal{C}^* = \arg\max_{\mathbf{V},\mathcal{C}} D\left(\phi(\mathbf{V}, \mathcal{C}; J, \{M_j\}, \{N_j\}, K)\right) \quad (3)$$

where D is a design criterion and ϕ is the SCMA system as it was described in Subsection III-A. However, when a regular SCMA system is fully loaded, one mapping matrix solution is possible and hence it is automatically the optimal one.

Existing SCMA codebooks designs simplified this optimization problem into a suboptimal multi-stage approach [14], such that the design of SCMA codebook is performed in three main steps: firstly a mother constellation, $\boldsymbol{C}_m$, is designed, then a user-specific transformation matrices, $\mathbf{T}_j$, are used to generate user-specific multi-dimensional constellations which are finally spread to generate the J codebooks. Hence, the optimization problem in (3) is reformulated as follows,

$$\{\mathbf{T}_j^*\}, \boldsymbol{C}_m^* = \arg\max_{\{\mathbf{T}_j\}, \boldsymbol{C}_m} D(\phi(\mathbf{V}^*, \{\mathbf{T}_j\boldsymbol{C}_m\}; J, \{M_j\}, \{N_j\}, K)) \quad (4)$$

such that the j^{th} codebook is calculated by,

$$\boldsymbol{C}_j = \mathbf{V}_j^*\mathbf{T}_j^*\boldsymbol{C}_m^*. \quad (5)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

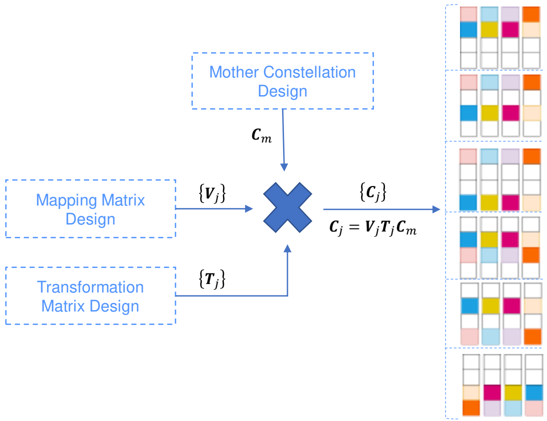

Figure 9. A block diagram of SCMA codebook design procedure.

This codebook design procedure is illustrated in Figure 9. In the following parts of this section, a review of existing designs of mother constellation and transformation operators will be introduced.

A. Mother Constellation Design

The mother constellation matrix of regular SCMA consists of N rows (or dimensions) of size M which allows to represent each $\log_2(M)$ bits with a codeword of N entries, the i^{th} entry is a complex value which belongs to the i^{th} dimension of the multi-dimensional constellation. The designed codebook must possess a good distancing properties among the points of the overall multi-dimensional constellation according to the criterion D such that performance can be enhanced. Exiting design criteria, D, including Euclidean distance, Euclidean kissing number, product distance, modulation diversity order and others are reviewed in [14]. These criteria are usually exploited to evaluate the mother constellation, in this case the applied transformations which are used to generate the J codebooks must be designed to preserve the characteristics of the mother constellation.

Hence, having several dimensions gives SCMA additional degrees of freedom, this results in an inherent shaping gain which is defined by the ratio of the minimum distance between the codewords of normalized multi-dimensional constellation of size M to the minimum distance between the points of a normalized traditional one-dimensional constellation of the same size. For example, the shaping gain of 4-point two-dimensional constellation as proposed in [64], [75] over quadrature phase shift keying constellation, in terms of Euclidean distance, is up to 1.25 dB. The gain on the minimum distance is translated in terms of better system performance.

With the aim to optimize the design of the mother constellation, $\boldsymbol{C}_m$, several works was proposed. For instance, in [75] the M-Beko mother constellation can be generated by minimizing the average alphabet energy for a given minimum Euclidean distance between any two constellation points. Similarly, the M-Peng scheme [76] is designed based on maximizing the minimum Euclidean distance between constellation points of different users. In [77], a complex constellation building is proposed. Such that its imaginary part is independent of its real one, which can help to reduce the decoding complexity, a shuffling operation can be employed to separate the imaginary and the real part of the complex constellation.

One approach to further reduce the complexity of the receiver is to conceive a low-projection mother constellation [77]–[81], that is to employ a lesser number of colliding constellation points over each dimension, however the codewords of each user can still be decoded since they are distinct on other dimensions as shown in Figure 10. This will decrease the complexity at the detector, for instance, MPA complexity can be reduced from M^{d_f} to $M_p^{d_f}$, where M_p is the size of the constellation after projection. The best employed criterion to design the low-projection constellations is *product distance* which has to be adjusted to be as low as possible without degrading the performance in the high signal-to-noise ratio (SNR) zone.

However, optimizing spreading codes and constellations is not a new problem, different approaches with their associated criteria exist in the literature [82]–[87]. These works inspired a variety of multi-dimensional SCMA constellation designs. Based on the quadrature amplitude modulation (QAM), the TMQAM scheme was proposed in [79] where the design of the proposed constellation uses a shuffling method which establishes the N-dimensional complex constellation from the Cartesian product of two N-dimensional real symbols with a specific Euclidean distance. Then, a rotation is applied to maximize the minimum product distance of both N-dimensional constellations. Similar to TMQAM, the MLQAM [80] technique applies the shuffling method but its constellation has a low number of projections. Also, the authors in [88] proposed a constellation for SCMA systems over Rayleigh fading channels based on a criterion derived from cutoff rate of MIMO systems, their proposition was denoted as M-Bao. A spherical coding was also proposed in [89] by the same authors of [88], however they shown that M-Bao codebooks outperform the spherical coding based ones. The TMQAM, MLQAM and M-Bao are all based on Cartesian product of two $\log_2(M)$-QAM which constitutes the M corners of a $\log_2(M)$-dimensional hyper-cube. The authors in [90] proposed the MHQAM scheme based on an optimization of rotation angles of the hyper-cube, this leads to a reduction of MPA complexity from M^{d_f} to $(\log_2(M))^{d_f}$. For a better illustration of mother constellation design, two examples of 2-dimensional ones with 4-codewords are provided in Table II and Figure 11.

Some multi-dimensional constellations are constructed by exploiting multi-radius rings [80], [92]–[94], i.e. the constellation points of a given dimension are not randomly placed but they belong to concentric rings. The M-point circular constellation MCQAM [80] is based on the analysis of the signal space diversity for MIMO systems over Rayleigh fading channels with a low number of projections for each complex dimension. MPA complexity is then reduced from M^{d_f} to $(M-1)^{d_f}$. A multi-dimensional SCMA codebook design based on star-QAM constellations was proposed for uplink SCMA systems in [92], [93]. That is the first dimension of

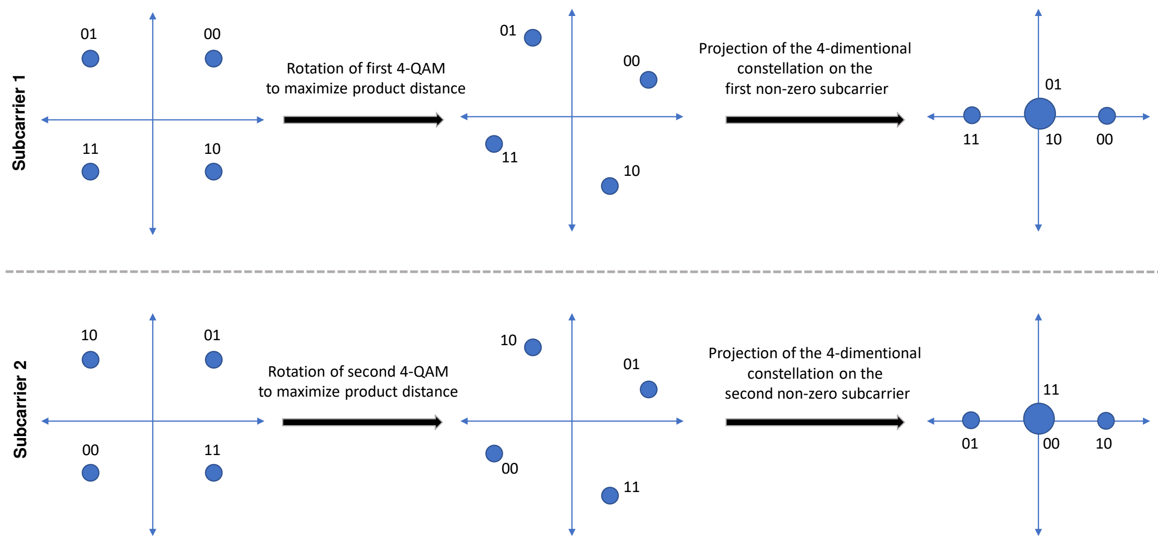

Figure 10. Low-projection constellation : an example of QAM SCMA codeword of size $M = 4$ with two non-zero resource elements.

Table II
TWO EXAMPLES OF 2-DIMENSIONAL CONSTELLATIONS WITH 4-CODEWORDS
WHERE $x_{m,k}$ IS THE k^{TH} ENTRY OF THE m^{TH} CODEWORD m, THAT IS $x_{m,k}$ BELONGS TO DIMENSION k.

Codeword m	T4QAM		4LQAM	
	$x_{m,1}$	$x_{m,2}$	$x_{m,1}$	$x_{m,2}$
1 (00)	$+\frac{3}{\sqrt{10}}$	$+\frac{1}{\sqrt{10}}$	$-\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}i$
2 (01)	$-\frac{1}{\sqrt{10}}$	$+\frac{3}{\sqrt{10}}$	$-\frac{\sqrt{2}}{2}$	$+\frac{\sqrt{2}}{2}i$
3 (10)	$+\frac{1}{\sqrt{10}}$	$-\frac{3}{\sqrt{10}}$	$+\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}i$
4 (11)	$-\frac{3}{\sqrt{10}}$	$-\frac{1}{\sqrt{10}}$	$+\frac{\sqrt{2}}{2}$	$+\frac{\sqrt{2}}{2}i$

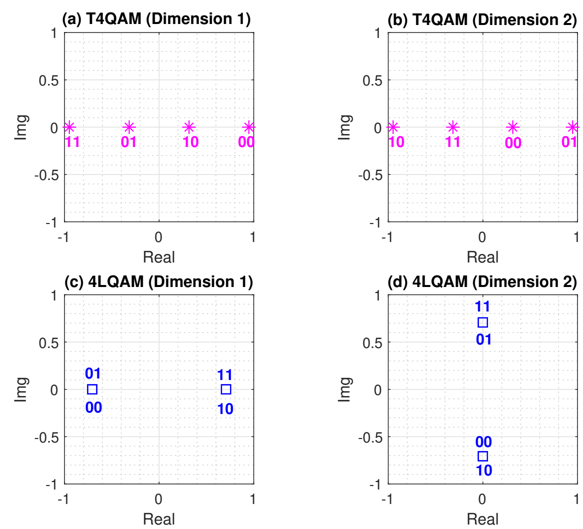

Figure 11. Illustration of two examples of 2-dimensional constellations with 4-codewords, i.e. T4QAM [79] and 4LQAM [80]: (a) projection of T4QAM over first dimension, (b) projection of T4QAM over second dimension, (c) projection of 4LQAM over first dimension and (d) projection of 4LQAM over second dimension.

mother constellation is constructed using a M-dimensional star-QAM constellation, then, the other dimensions are obtained by scaling and permuting the points of the first one, an example is illustrated in Figure 12. The mother constellation parameters are obtained through computer search inspired by the approaches in [82], [86]. This approach is powerful for designing codebooks with large size and/or high dimension. The optimization objective in [93] was to minimize the pairwise error probability between two transmitted codewords $\mathbf{x}_a, \mathbf{x}_b$ which is given by,

$$\mathbb{P}(\mathbf{x}_a, \mathbf{x}_b|\mathbf{H}) = Q\left(\sqrt{\frac{\|\mathbf{H}\left(\mathbf{x}_a - \mathbf{x}_b\right)\|^2}{2N_0}}\right). \quad (6)$$

This criterion can be interesting since it is applied directly on the generated codewords of the J codebooks and not on the mother constellation.

Multi-stage optimization of another ring-based approach was proposed in [94]. Each ring is composed of uniformly spaced phase shift keying (PSK) points such that the whole rings form an amplitude and phase shift keying (APSK) constellation which is capable of outperforming the classic square shaped QAM constellation in peak-power-limited systems. The first dimension is designed by maximizing the coded modulation capacity, then the other dimensions of the mother constellation are deducted using optimized permutations.

The optimization problem defined in (3) and (4) is usually non-convex quadratically constrained quadratic programming problem [75], [76]. This class of optimization problems is NP-hard, and it is generally difficult to solve. This problem can be reformulated/linearized and consequently transformed into a convex second-order cone programming one, hence it can be solved efficiently, e.g., by CVX [95]. This procedure is known as the convex-concave method [75]. Another way to solve this type of problems is to approximate it based on a semidefinite

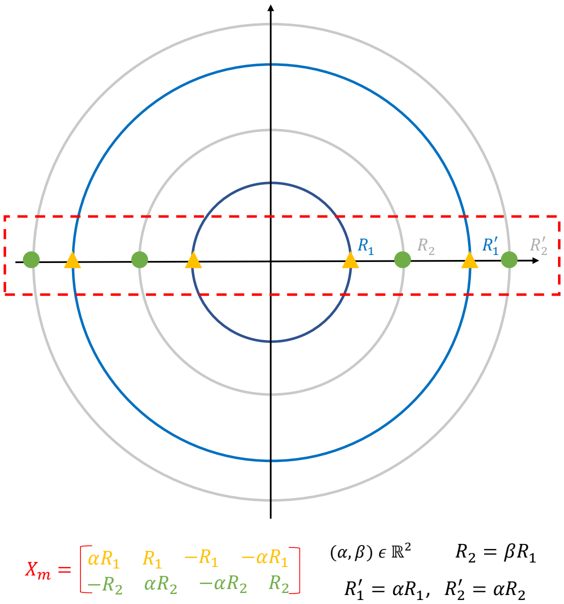

Figure 12. An example of four-rings star-QAM mother constellation for the design of a SCMA codebook of size $M = 4$ and sparsity degree $N = 2$ where α and β are 2 reel design parameters.

relaxation, then the original solution can be extracted from the optimal solution of the approximated problem by Gaussian randomization approach [76]. However, in most exiting works, the optimization problem is reformulated and then solved by an exhaustive numerical computer search [88], [90], [93], [94], [96]. This approach could be computationally complex, some solutions were proposed to overcome this difficulty. The authors in [88], [93] proposed a compact reduced parameterization of matrices to be optimized which makes this simplified search process feasible, for instance, the whole design process in [93] is based on two parameters. In [94], the complexity of search can be reduced by iteratively searching inside a randomly selected list or by using the binary switching algorithm which penalizes the symbols with the strongest contribution to a *bad* performance. In [90], the research zone of rotation angles is first reduced based on *maximum a posteriori* analysis, then second-order rational polynomial is used to find a close-to-optimal angle based on the Levenberg-Marquardt algorithm which is efficient for non-linear least squares curve-fitting problems, this operation has a much lower complexity than an exhaustive search [97].

B. Transformation Operators Design

Once having the mother constellation of N-dimension, $\boldsymbol{C}_m$, different sets of operators can be applied on top of it to build multiple sparse codebooks for several layers of SCMA. As one of the fundamental steps in the design of SCMA codebooks, the transformation operators design was investigated in recent researches works [93], [94], [98]–[102]. A transformation operator, $\mathbf{T}_j$ (see equation (5)), could be either one or a combination of typical operators such as complex conjugate, rotation operator, interleaving and vector permutation.

In most works, preserving an adequate Euclidean distance profile of the mother constellation is based on applying a unitary rotation matrix [64], [75]–[77], [79], [91], [103]–[105]. The authors in these research works have combined the mapping matrix and the transformation one to form a new rectangular matrix, $\mathbf{F}_T$. An example, for $K = 4, N = 2$ and $M = 4$, is given by [91],

$$\mathbf{F}_T = \begin{bmatrix} 0 & \varphi_1 & \varphi_2 & 0 & \varphi_3 & 0 \\ \varphi_2 & 0 & \varphi_3 & 0 & 0 & \varphi_1 \\ 0 & \varphi_2 & 0 & \varphi_1 & 0 & \varphi_3 \\ \varphi_1 & 0 & 0 & \varphi_3 & \varphi_2 & 0 \end{bmatrix} \tag{7}$$

where $\varphi_1 = e^{j\theta_1}, \varphi_2 = e^{j\theta_2}$ and $\varphi_3 = e^{j\theta_3}$. Traditionally, $\theta_1 = 0, \theta_2 = \frac{\pi}{3}$, and $\theta_3 = \frac{2\pi}{3}$. In this circumstance, the codebook of user 1, for instance, is calculated based on the following mapping and transformation matrices,

$$\mathbf{V}_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{T}_1 = \begin{bmatrix} e^{j\theta_2} & 0 \\ 0 & e^{j\theta_1} \end{bmatrix}.$$

The matrix $\mathbf{F}_T$ respects the Latin criterion where not only non-zero elements in each row are distinct but also those in each column which leads to controlling both dimensional dependency and power variation of the multi-dimensional constellation, while keeping the Euclidean distance unchanged [77], [103]. Figure 13 shows the normalized 2-dimensional constellations with 4-codewords, generated by applying unitary rotation transformation, as described in equation (7), on the T4QAM mother constellation. The projections of the constellation of each user on its associated two REs are depicted in subfigures. The complete codebooks, as introduced in [91], are given in appendix A.

In [88], multi-user codebooks are obtained via specific computer-designed rotation matrices instead of the unitary rotation ones. Furthermore, a two-dimensional specific user rotation is applied to the mother constellation proposed in [80] to generate the different SCMA codebooks. In [100], the SCMA design and the security of the link are combined, the codebooks are generated by rotating the mother constellation with random angles extracted from channel phases, this requires to know the CSI. These encrypted codebooks protect exchanged information with low complexity.

An optimization of transformation operator was proposed in [99] based on a novel criterion to select the most appropriate permutation set in order to improve the probability of reliable detection of the first user which largely improves the performance of MPA detector. The proposed criterion tries to maximize the sum of distances among dimensions of interfering codewords multiplexed on each RE. A permutation-based SCMA scheme is proposed in [106], the idea is that the codebook of user j is designed by mapping the encoded codeword to N among K REs whose positions are defined according to the values of data bits to be transmitted, i.e. different non-zero locations of encoded complex vectors are assigned to different codewords. This approach is different from the majority of SCMA designs where the fixed positions of non-zero elements of each user are determined by the columns of the factor graph

Table III
REVIEW OF EXISTING SCMA CODEBOOK DESIGNS

Reference	Year	Approach	Uplink / downlink	Optimization criterion	Description
Beko *et al.* [75]	2012	M-Beko	Uplink & downlink	Average alphabet energy for a given minimum Euclidean distance	M-Beko method formulates the constellation design as a non-convex optimization problem which objective is to minimize average alphabet energy between any two constellations points. This method can offer an optimal solution.
Boroujeni *et al.* [79]	2013	TMQAM	Uplink & downlink	Euclidean distance, product distance & average product distance	Using a unitary rotation to produce a multi-dimensional mother constellation. This document introduces also the first low-projection multi-dimensional constellation. USA Patent US9509379B2.
Taherzadeh *et al.* [77]	2014	Shuffling QAM	Uplink & downlink	Euclidean distance & product distance	A systematic approach based on lattice constellations is employed to design multi-dimensional constellation with a reasonable minimum Euclidean distance which is then rotated to reach a good product distance. Shuffling operation can be applied.
Yu *et al.* [92] [93]	2015 2018	Star-QAM	Uplink	Pairwise error probability	Codebook with large minimum Euclidean distance, using star-QAM improves SCMA BER gain without sacrificing the system implementation complexity.
Zhang *et al.* [70]	2015	IrSCMA	Uplink	Not specified	Irregular SCMA: Various sparsity degrees for various user needs. Authors introduced the notion of degree distribution of SCMA. The first proposition of an irregular SCMA.
Yu *et al.* [71]	2016	IrSCMA	Downlink	EXIT chart	Irregular SCMA: this paper extends the proposition in [70] via an optimization of the employed rotation angles. The degree of user superposition on a each RE is constant in contrast to [70].
Bao *et al.* [88]	2016	M-Bao	Uplink	Cutoff rate of MIMO system	Rotation matrices are obtained through exhaustive computer search over compact parameterizations of unitary matrices, a suboptimal solution was applied. Perfect channel knowledge is assumed.
Bao *et al.* [89]	2016	Spherical codes	Uplink	Squared Euclidean distance	Spherical codes design is formulated as a non-convex second-order cone programming problem. Hybrid lattice and spherical codes based constellations are also discussed. Low peak-to-average power ratio. -
Metkarunchit [80]	2017	Low-projection MC	Downlink	Cutoff rate of MIMO system	Circular QAM for SCMA mother constellation design, complexity of MPA depends on the reduced number of constellation points.
Peng *et al.* [76]	2017	M-Peng	Uplink & downlink	Euclidean distance	A joint design of codebook and mapping matrices. The proposed solution is semi-definite relaxation of non-convex quadratically constrained quadratic programming.
Yan *et al.* [99]	2017	Dimension distance-based design	Uplink & downlink	Sum of distance between dimensions of interfering codewords	Codebook design is based on turbo trellis coded modulation. Phase rotation and interleaving employ an appropriate permutation set which is selected according to the introduced criterion.
Lai *et al.* [100]	2017	Dynamic codebook design	Downlink	Not specified	A transmission of SCMA codewords designed using random angles extracted from CSI wich will grant more secure transmission process. Upper-bound-aided codebook design was introduced.
Zhai *et al.* [72]	2017	IrSCMA, QAM	Uplink	Not specified	Flexible resources scheduling scheme according to user's features.
Bao *et al.* [94]	2018	Low-projection mother constellation	Uplink	Coded modulation capacity	Multi-stage optimization of multi-radius rings based multi-dimensional constellation by permuting an one-dimensional constellation. Introduction of a new design criterion. Application of bit-interleaved BICM with iterative multiuser detection.
Vameghesta-hbanati *et al.* [90]	2020	MHQAM	Uplink	Frame-error-rate of a LDPC-coded system	Construction of gray-labelled mother constellation based on hypercubes which is used along with BICM. Exhaustive search is employed.

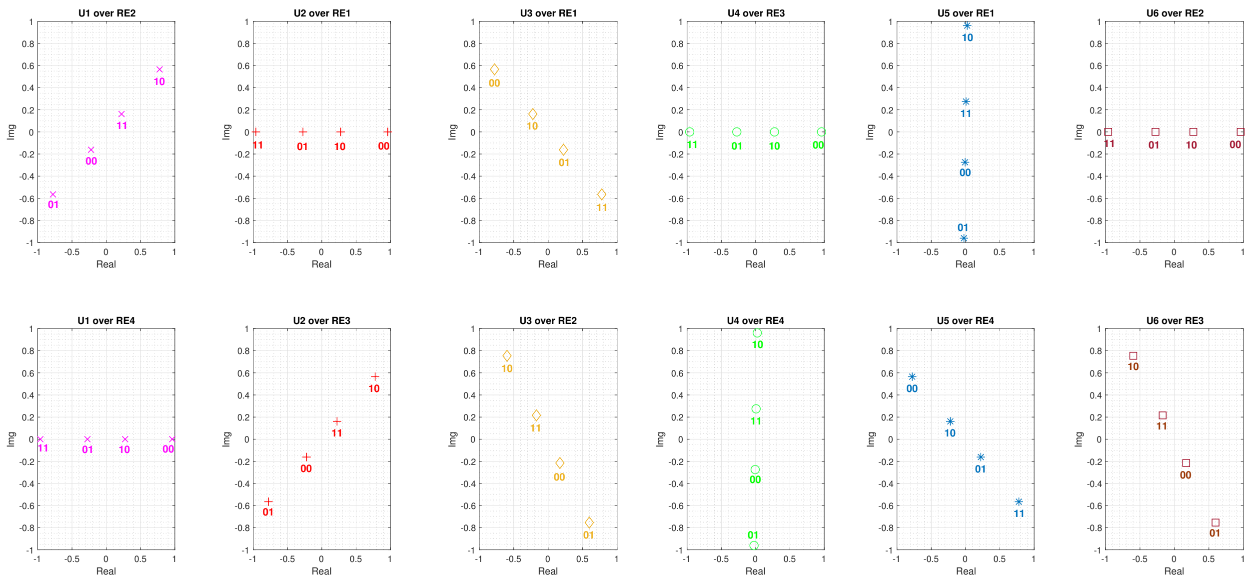

Figure 13. Illustration of the six 2-dimensional constellations with 4-codewords, generated by applying unitary rotation transformation, as described in equation (7), on the T4QAM mother constellation [91]: the projections of the constellation of each user on its associated two REs are depicted.

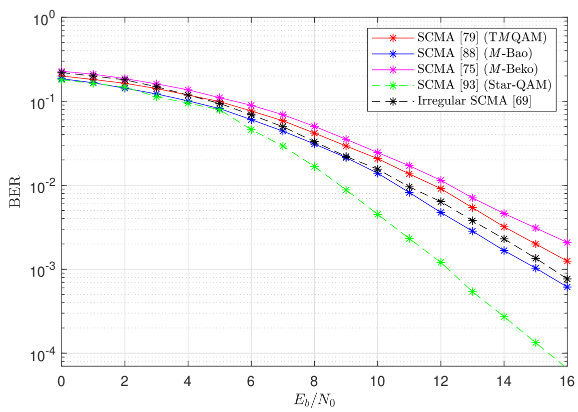

Figure 14. Performance evaluation of uplink SCMA system with different codebook designs through Rayleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6.

matrix, $\mathbf{F}$, as illustrated in Subsection III-A. The proposed permutation approach increases the spectral efficiency without a complexity overhead when compared to traditional SCMA. A combination of matrix permutation and rotation is employed as transformation operator in [92], [93].

It is worth mentioning that labeling affects the performance of SCMA system in terms of BER, hence it is important to employ the appropriate labeling. For instance, once the multi-dimensional constellation is designed in [94], its labeling is optimized such that the slope of the extrinsic information transfer (EXIT) chart is well adjusted.

A performance evaluation of uplink SCMA system with different codebooks through Rayleigh fading channel is shown in Figure 14, it is obvious that the design of codebook has an important effect on the system performance. A summary of existing SCMA codebook designs is presented in Table III. Furthermore, a review of SCMA multi-dimensional constellations design was proposed in [14] in which further details on some of the aforementioned SCMA codebooks can be found.

V. SCMA Detector Design

In this section, our objective is to highlight existing algorithms and mechanisms employed to enable the receiver to separate the superimposed codewords which were selected from one of the above presented SCMA codebooks. Each RE of the SCMA systems is occupied by several users which results in smaller Euclidean distances among the constellation points and could degrade the link performance. A block diagram of classic SCMA system is shown in Figure 15, after the OFDM demodulator, the transmitted signals must be estimated which requires to have some knowledge of the communication channel, to segregate the SCMA codewords and finally to decode the channel encoding after a deinterleaving operation. Most existing methods propose to sequentially apply the aforementioned steps, however some research works aim either to jointly decode SCMA and channel correction codes, or to design joint detection methods for MIMO-SCMA systems. A comprehensive review of stand-alone or joint SCMA detectors is provided in the following subsections. Also, Table V summarizes SCMA detection techniques by showing their approach, their assumptions and their complexity order.

Most existing works on SCMA detectors employ MPA or one of its variations, or a combination of MPA and other methods. Hence, it is straight-forward to start by presenting MPA in the next subsection before reviewing other existing techniques. However, Figure 16 shows the structure of this section along with the most important detection methods.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

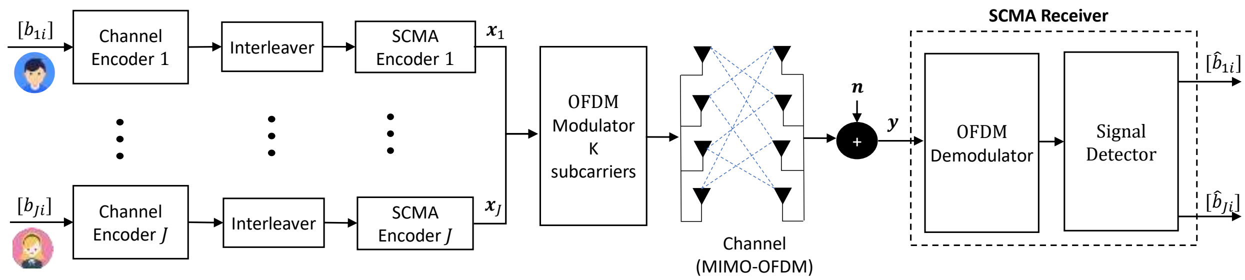

Figure 15. Block diagram of SCMA system model.

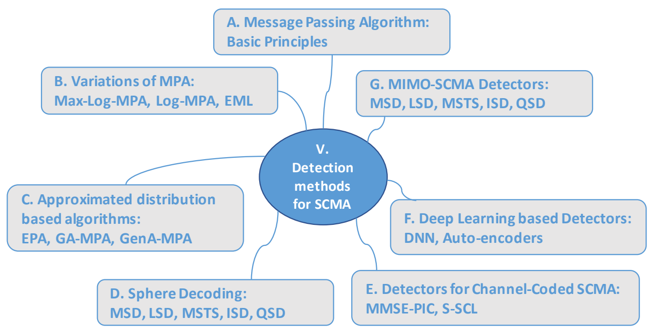

Figure 16. Detection methods for SCMA systems.

A. Message Passing Algorithm

MPA is defined as an iterative parallel decoding technique based on passing the extrinsic information from function nodes (FNs) to variables nodes (VNs) and vice versa [37], [107] as shown in Figure 7. In each iteration, each FN of the factor graph computes its outgoing message to a given VN depending on the incoming messages received from the rest of VNs. Then, each VN will send back its message depending on the received messages from the reminder of FNs. Finally, one iteration is considered as completed when one outgoing message has passed in both directions along every edge. The log-likelihood-rates (LLRs) of each coded bit are calculated after few iterations in order to estimate the bites of each user j.

The MPA is based on three main steps as shown in Algorithm 1 for a SCMA system of J users using M constellation points through K orthogonal REs.

B. Variations of MPA

A complexity evaluation of MPA reveals that it relies on a large number of exponential calculus which are of high complexity. A mathematical simplification is to approximate the logarithm of a sum of exponential operations into a maximum operation which is very less expensive in term of complexity. This lead to introduce a simplified version of MPA denoted Max-Log-MPA [108], [109]. To mitigate the performance degradation resulting from the aforementioned approximation, a correction term can be added which introduced another simplified version of MPA known as Log-MPA [109], [110]. Figure 17 depicts the BER obtained by the three previously mentioned variations of MPA, namely MPA, Log-MPA and Max-Log-MPA, through Rayleigh fading channel. It is obvious that, contrary to Max-Log-MPA, Log-MPA can achieve near-optimum performance when compared to MPA due to the aforementioned correction term which compensates for the performance loss of Max-Log-MPA. Furthermore, the performances of the three methods converge when SNR increases. However, the computational complexity of Log-MPA is still a big challenge for some practical implementation especially for energy-sensitive user equipment's in the downlink scenario. As the computation complexity of the above-mentioned detection algorithms increases exponentially with d_f, the detection still takes considerable time. The parameter d_f must be designed to be very small, which largely limits the choice of codebooks.

Also, in [111], a framework for an improved MPA based SCMA multi-user detector was designed. Two aspects of MPA are simplified and optimized. A lookup table scheme was introduced first to substitute the information calculation in the FN which will reduce the complexity and can guarantee a stable convergence of the MPA based detector. The authors proposed also two novel scheduling methods namely the single scheduling MPA (SS-MPA) and the multiple scheduling MPA (MS-MPA). For SS-MPA, the soft messages in the FNs and VNs are serially calculated and synchronously updated. On the other hand, MS-MPA employs multiple detectors to calculate in parallel and update the node messages by different orders. Both the SS-MPA and MS-MPA are optimized to converge more quickly than the conventional MPA.

A mechanism based on joint SIC-MPA was proposed in [112], the soft information at the SIC output is sent to the MPA. CS based techniques can be invoked to enhance the MPA based detector as studied in [113], the proposed method is referred to as a CS-aided MPA detector. The detector is divided into two phases: an initial detection is performed by few MPA iterations, this operation is followed by sparse error-correction based on the fact that the symbol error produced by initial detection tends to be sparse. Hence, a CS-based greedy pursuit algorithm is applied to estimate the error accurately, and consequently the initial estimated vector is updated by adding the estimated error vector.

Authors in [114] proposed serial MPA algorithm based on fairness where, instead of calculating the messages in parallel, the probability information of a given user output codeword is updated immediately as soon as the information of one of its corresponding FNs is updated. This method can improve the convergence performance of the algorithm and reduce the computational complexity. An edgewise serial based MPA was

Algorithm 1: Message Passing Algorithm

Input: $\mathbf{y}, N_0, \boldsymbol{C}_j, \mathbf{h}_j, j = 1, \cdots, J$.

Result: Estimate the value of coded bits for each user j, $j = 1, \cdots, J$.

Definitions

Variables nodes (VNs) represent users,

Function nodes (FNs) represent resources or subcarriers,

$\mathcal{U}(k) = \{j, 1 \leq j \leq J \mid \text{VN}_j \text{ is connected to FN}_k\}$,

$\mathcal{R}(j) = \{k, 1 \leq k \leq K \mid \text{FN}_k \text{ is connected to VN}_j\}$.

Step1: Initialization

The prior probability for each codeword is given by: $V^0_{j\to k}(\mathbf{x}_m) = \mathbb{P}(\mathbf{x}_m) = \frac{1}{M}, j = 1, \cdots, J, k \in \mathcal{R}(j)$

Step2: Iterative message passing along edges

while N_{iter} **do**

1) FN update: the message to be passed from FN_k to one of its neighbors VN_j, $k = 1, \cdots, K, j \in \mathcal{U}(k)$, for a given codeword $\mathbf{x}_m \in \boldsymbol{C}_j, m = 1, \cdots, M$, is calculated as,

$$U^t_{k\to j}(\mathbf{x}_m) = \sum_{\mathbf{x}_i|i\in\mathcal{U}(k)\backslash j} \exp\left\{-\frac{1}{N_0}\left\|y_k - \sum_j h_{j,k}x_{m,k}\right\|^2\right\} \prod_{i\in\mathcal{U}(k)\backslash j} V^{t-1}_{i\to k}(\mathbf{x}_m)$$

2) VN update: the message to be passed from VN_j to one of its neighbors FN_k, $j = 1, \cdots, J, k \in \mathcal{R}(j)$, for a given codeword $\mathbf{x}_m \in \boldsymbol{C}_j, m = 1, \cdots, M$, is given by,

$$V^t_{j\to k}(\mathbf{x}_m) = \frac{\prod_{i\in\mathcal{R}(j)\backslash k} U^{t-1}_{i\to j}(\mathbf{x}_m)}{\sum_{\mathbf{x}_l\in\boldsymbol{C}_j}\prod_{i\in\mathcal{R}(j)\backslash k} U^{t-1}_{i\to j}(\mathbf{x}_l)}.$$

Normalization is necessary to keep the algorithm numerically stable.

end

Step 3 : Taking a decision (LLR at each VN)

1) For each layer j, $j = 1, \cdots, J$, the a posteriori probability of codeword $\mathbf{x}_m, m = 1, \cdots, M$ is defined as,

$$\mathbb{P}(\mathbf{x}_m) = \prod_{k\in\mathcal{R}(j)} U^{N_{\text{iter}}}_{k\to j}(\mathbf{x}_m)$$

2) Log-Likelihood-Rate for each coded bit, $b_i, 1 \leq i \leq \log_2(M)$, is represented by,

$$\text{LLR}(b_i) = \log\left(\frac{\mathbb{P}(b_i = 0)}{\mathbb{P}(b_i = 1)}\right) = \log\left(\frac{\sum_{\{\mathbf{x}_m\in\boldsymbol{C}_j|b_i=0\}}\mathbb{P}(\mathbf{x}_m)}{\sum_{\{\mathbf{x}_m\in\boldsymbol{C}_j|b_i=1\}}\mathbb{P}(\mathbf{x}_m)}\right)$$

Finally, each bit is decided according to the value of its LLR as following,

$$\hat{b}_i = \begin{cases} 1 \text{ if LLR}(b_i) \leq 0 \\ 0 \text{ otherwise.} \end{cases}$$

proposed in [115], the idea is to update the messages in an edge by edge way which allows the messages to the same VN to be updated at different times such that a more reliable extrinsic information is used in a balanced manner.

Several SCMA detection methods employ a threshold in order to enhance the performance and/or reduce the complexity [116]–[119]. A threshold-based MPA was introduced in [116], it is based on comparing the LLR of each coding bit with a given threshold in order to evaluate its reliability such that its value can be decoded in advance and the updating of the user can be stopped. However, this may cause the loss of the posteriori soft information of other users who occupy the same RE and results in a reduction in the accuracy of likelihoods calculation. The authors in [117] employed the less complex Max-Log-MPA instead of MPA, and they chose to stop updating a VN based on a combination of the codeword reliability and the *user node stability*. A node is considered stable when the more reliable element in its codeword is still the same during several iterations. A selection criterion was introduced in [118] in order to reduce the computational complexity by allowing only the connections between a FN and a VN with belief ratio lower than a given threshold in the last iteration to interfere in the message passing process at the current iteration. In [119], a threshold is employed to choose the edges which can interfere in calculating each MPA message. The expressions of the probability of selecting a given number of edges and that of the average complexity order were derived as a function of the threshold. A belief judgment criterion is applied in [120] on each FN and VN to judge the early termination, i.e. to decide when the convergence is achieved, the authors proposed also a stability judgement based self-adaption algorithm to predict and update the belief of next

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

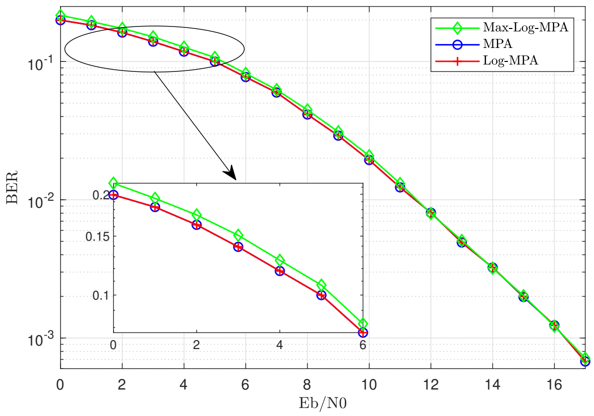

Figure 17. Performance comparison of MPA, Log-MPA and MAX-Log-MPA variations through Rayleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6.

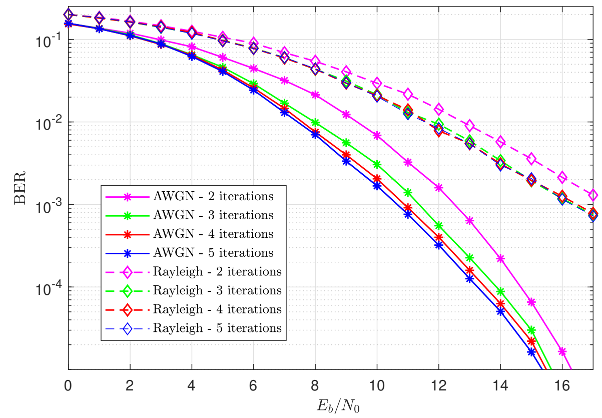

Figure 18. MPA performance comparison for different number of iterations through AWGN channel and Raleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6.

iteration based on the convergence tendency.

A partial marginalization MPA (PM-MPA) was introduced in [121], the codewords of a given number of users are determined after a predefined number of iterations, the concerned users are chosen randomly or based on the codewords reliability. Only the message of the undetermined users will be updated with the determined codewords in the remaining iterations. After the maximum number of iterations, the codewords of all users will be finally determined. In [122], [123], the authors introduced a SCMA trellis representation after mapping SCMA constellation to a Galois field. Based on the introduced trellis representation, a low-complexity truncated-messages based decoding method was proposed. The principle of truncation means selecting the most reliable messages (and their corresponding symbols) after their calculation at each user node. This method was denoted as extended max-log (EML). They further propose a channel adaptive EML algorithm to truncate the messages with a criterion that can be adapted according to the CSI.

Most MPA variations are based on a fixed number of iterations which is not optimal since a large number of iterations will considerably increase the complexity, while a low one will lead to less performing detectors. Figure 18 evaluate the performance of MPA, in terms of BER, for different number of message passing iterations when the channel is assumed to be AWGN or Rayleigh distributed. In the two cases, the performance improves as the number of iterations increases, nevertheless no significant improvement can be observed beyond a certain limit. Moreover, the authors in [109] affirmed that the same convergence behavior can be observed for Log-MPA and Max-Log-MPA. Therefore, a reasonable number of iterations can be set to 4. A flexible number of iterations was allowed in [124] by supervising the convergence rate of each codeword probability which will guarantee to reduce the complexity without compromising the performance.

C. Approximated distribution based algorithms

Gaussian-approximated (GA)-MPA was applied for CDMA and massive MIMO systems [125]. In the last few years, several research works dealt with how approximated distribution based MPA can be used to detect SCMA symbols [126]–[129]. One idea is that the message update operation at FNs is simplified by updating the symbol likelihoods for each one among its connected VNs based on the fact that the other VNs' messages are approximated as a Gaussian interference. In [126], two simplified algorithms were proposed, namely generalized approximate MPA (GenA-MPA) and SIC-GenA-MPA. The approximation adopted in GenA-MPA transforms the vector estimation problem into a scalar one, hence the complexity is reduced and hardware solutions become feasible. The SIC-GenA-MPA method further simplifies the GenA-MPA one by calculating the closest value to the user codewords, and then eliminating the calculated user layer of MPA based on the generalized approximate result. In [127], maximum likelihood (ML) is employed for a selection of FNs while GA-MPA is performed on the remaining ones. The authors in [128] combined matched filtering with iterative GA based detector in an uplink SCMA system. In order to alleviate the absence of CSI, a training phase is performed to estimate the matcheded filtered channel before data transmission. The training overhead can be neglected, and the complexity of this method is of polynomial order. The factor graph was extended for MIMO-SCMA systems in [129] such that two GA based detectors were proposed. The complexity was pulled down from exponential to linear order.

Another approach to design SCMA detectors is to estimate posterior distributions through distribution projection [130]–[133]. The complexity of this approximate Bayesian inference method depends on the simplicity of the distribution to which we are projecting. In [130], posterior distributions are projected to a complex Gaussian distribution such that the message passing is reduced to mean and variance calculation. In other words, only expectations are retained to be iterated, this method is denoted EPA [134]. EPA reduces the complexity order from exponential to linear as presented in Table V. Simulation results showed that EPA receiver achieves nearly the performance of a traditional MPA receiver with less complexity [130]. The authors in [132] applied the Monte Carlo Markov chain method to study the SCMA detection problem, a new joint update parallel probability sampler was introduced. EPA based detectors for MIMO-SCMA systems were proposed in [131], [133].

The BP can be used to calculate the posterior probability and consequently to iteratively detect SCMA signals. For instance, BP based detector was proposed in [135] for uplink SCMA systems. The authors studied the asynchronous transmission case where signals from different users suffer from different delays at the receiver, consequently new interferences caused by the delay must be taken into consideration. Hence, decoding is only possible when the entire signal sequence is received. A Tanner graph illustrated the corresponding BP algorithm before calculating the conditional probability. Also, a joint factor graph representation of user activity detection, channel estimation and data detection was proposed in [136], then a BP based message passing algorithm is used after approximating its distributions by Gaussian ones with the minimized Kullback-Leibler divergence.

Bethe approximation based message passing detection algorithms were proposed for massive MIMO systems [137] and SCMA systems [138], [139]. The idea is to avoid the high-complexity direct marginalization of the posterior probability of users' symbols by finding its appropriate trail distribution which is approximated by minimizing Bethe free energy. The complexity of Bethe approximation based algorithm only increases linearly with the number of users. To further improve the performance, a damping scheme was employed.

The discretized message passing algorithm (DMPA) was proposed in [140], [141]. The idea is to reduce the size of the joint constellation on each FN by discretizing probability density functions (PDFs) in the VNs. Actually, discretization means approximating a given value by the nearest sampling point where points are chosen according to a given sampling interval. The message update is reformulated as a convolution which is calculated by 2-dimensional fast Fourier transform (FFT). This procedure will reduce the complexity of the exhaustive search on the signals from exponential to polynomial.

D. Sphere Decoding

One way to reduce the complexity of the SCMA maximum likelihood detector is to restrict the space of codeword research to a hyper-sphere centered around the received signal vector as depicted in Figure 19, this technique is called sphere decoding (SD) [142]. Several research works proposed how to apply SD to SCMA detector [81], [96], [143]–[151] in the aim of reducing the complexity while still providing good performances.

A region-restricted detector with an improved Log-MPA is proposed in [144]. The complexity is reduced by avoiding unnecessary calculations when searching the superposition constellation exhaustively, the proposed detector updates the FNs only within a restricted search region while constellation

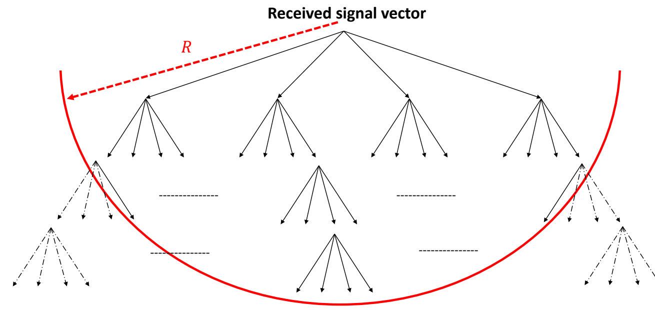

Figure 19. Graphical representation of sphere decoding. The objective is to reduce the space of superposed constellation point research to a hyper-sphere of radius R centered around the received signal vector.

points outside the search region are neglected. A SD-based MPA method proposed in [145] where the spherical search zone in adjusted according to the noise power at the receiver. However, SD-MPA suffers of high complexity in the low SNR region and a probable performance degradation in the high SNR zone. In [146] authors proposed a reduced-complexity optimal modified SD (MSD) detection scheme for SCMA. The proposed MSD achieves the performance of the optimal ML detection with an average complexity compared to MPA as shown in Table V. Another method to reduce the research space before implementing the MPA is the list sphere decoding (LSD) [81]. The authors exploited the fact that SCMA codewords are basically complex lattice constellation points to limit the research to a list of all the candidate lattice points within a given radius around the received vector. This list can be further reduced by eliminating some redundancy via some node pruning techniques.

The authors in [150] proposed a modified single tree search (MSTS) to obtain soft outputs for coded SCMA. The MSTS algorithm tries to avoid the unnecessary node searching by sorting channel matrix in ascending order, and by employing a non-zero low bound Schnorr-Euchner enumeration, before searching in order to reduce complexity during the search stage. MSTS not only has lower computational complexity than MPA based detectors but also keeps its performance level close to that of MPA. One drawback of the aforementioned methods is that they cannot support some of the non-constant modulus constellations, that is why a a new improved SD (ISD) method that can support any constellation topology was introduced in [151].

Quantum computing is an emerging technology which can be applied to solve some wireless communications problems, it can especially enhance the performance and/or reduce the complexity of some search problems [152]. In [96], two quantum search algorithm-aided SCMA decoding methods were proposed, namely, the quantum-assisted MPA (Q-MPA) and the quantum-assisted sphere decoder based MPA (QSD-MPA). The message updating in Q-MPA is accelerated by exploiting the results of a quantum search which is applied at the FNs. Then, the messages are updated in the two directions as the classical MPA. In regards to QSD-MPA, the MPA is preceded by a quantum search whose objective is to find all legitimate codeword combinations within a given hyper-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

sphere.

E. Detectors for Channel-Coded SCMA

Each wireless link requires an error-correction channel coding including turbo codes which was used in 4G systems, and low-density parity-check (LDPC) codes and polar codes which are adopted by the 5G NR standard [153], [154]. SCMA detectors which include the error-correction codes structure into their design began to gain more attention. For instance, the LDPC code was employed as the channel coding scheme for the SCMA system in [155], [156] while a polar-coded SCMA was studied in [157]–[159].

In [155], the sparse factor graphs of LDPC coding and SCMA are combined into a joint sparse graph in order to jointly perform the decoding and the detection on one graph. Then, MPA was simplified by using partial message passing based on a joint trellis representation. Based on their simulations, the authors showed that the joint approach is compatible with delay-sensitive applications since it reduces the processing latency compared with disjoint turbo approaches. Another approach was studied in [156] by optimizing the interface between the SCMA detector and the LDPC decoder which made a hardware solution feasible. The authors proposed a minimum mean-square error with parallel interference cancellation (MMSE-PIC) detection in the FN operations, and a bit LLR values for message passing. In order to minimize the hardware overhead and to reduce the processing latency, both the proposed algorithm and very large scale integration architecture are jointly designed.

In [157], a simple combination of MAP detector for SCMA and a soft-input soft-output successive cancellation for polar coding was studied. The architecture of the receiver was adjusted by re-encoding the soft information of the polar codeword. The reconstructed information is fed into the iterative SCMA detection procedure which results in an additional coding gain. A joint factor graph of SCMA detector and polar decoder is employed in [158], and hence a joint iterative message updating operation was introduced. The EXIT charts based analysis provided a more understating of how to optimize the polar-coded SCMA system. Especially, a weight factor was conceived and optimized to mitigate the effect of the correlation among the soft outputs of polar decoder.

It is not evident how to estimate the CSI for some practical mMTC applications where short packets are essentially exchanged. That is why it is interesting to propose a joint channel estimation and decoding for polar-coded SCMA. In [159], the joint detection and decoding scheme is based on traditional Max-Log-MPA detector for SCMA and soft-successive cancellation list (S-SCL) algorithm for polar decoding. The S-SCL provides the prior symbol probability for the SCMA detector while SCMA detector calculates the prior information of the polar decoder. The joint detection and decoding scheme is serialized with a frozen bit error rate (FBER) mapper such that the CSI is updated at each iteration. The iterative joint channel estimation and decoding scheme is initialized with a sparse Bayesian learning based channel estimation.

F. Deep Learning based Detectors

Recent advances in deep learning allowed its application in wide-ranging fields such as computer vision, healthcare, speech, etc. Deep learning contributed also to advances in some aspects of wireless communication systems as channel estimation, automatic modulation recognition, spectrum management, network traffic prediction [160].

By the end of 2018, a big interest in deep learning based NOMA schemes have been observed in the literature [161]–[168] with the aim of enhancing multiple access performance. Such approach can be reasonably considered, not only based on deep learning breakthrough, but also based on the improvement of the computational performance of wireless systems. Gui *et al.* proposed, in their work [161], a supervised long short-term memory (LSTM) based PD-NOMA where channel states of multiple users are automatically tracked by the LSTM network which can learn the environment via offline training. It was proven that the LSTM-based framework is very suitable for user activity and data detection of PD-NOMA systems. In [162], a two user PD-NOMA scenario was considered, a significant performance improvement of a deep neural network (DNN) receiver over a conventional SIC is provided. Also, a joint deep learning of the transmitter precoding and the receiver decoding for downlink MIMO-NOMA was investigated in [167]. In [168], the authors considered the application of deep learning techniques, namely fully connected neural network and convolutional neural network, for symbol-level multi-user detection in Welch bound equality spread based NOMA system, this NOMA scheme employs low-correlation spreading signatures. The proposed deep learning based solution matches almost the ML detector performance.

In this survey, we are more interested in how deep learning was applied to design SCMA detectors. Very recently, this problem attracted some attention [169]–[174]. The aim is to design a detector by offline training a DNN such that one shot online non-iterative decoding is performed with a relative low-complexity.

In [169], denoising autoencoders are used to jointly design the encoder and the detector as shown in Figure 20. Hence, one DNN generates the codebook automatically and must learn how to efficiently map symbols to a complex constellation, another fully-connected DNN decodes the received vector to detect the symbols. The training is based on an end-to-end objective function which must be minimized in order to minimize the BER. The training data is randomly generated with a given corruption level. A similar approach is also studied in [174]. However, the aforementioned detector can not be used with any given codebook. That is why the proposition in [171] studied only the detection problem independently from how codebook is designed and generated. A sparsely connected DNN was designed such that the propagation between two of its layers is calculated based on how messages are passed between FNs and VNs in traditional iterative MPA. The idea is to unfold the factor graph iterations into the network layers which results in sparse DNN. The simulated training data are generated dynamically to have a different noise level at each step. The same unfolding principle was studied in [170] in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

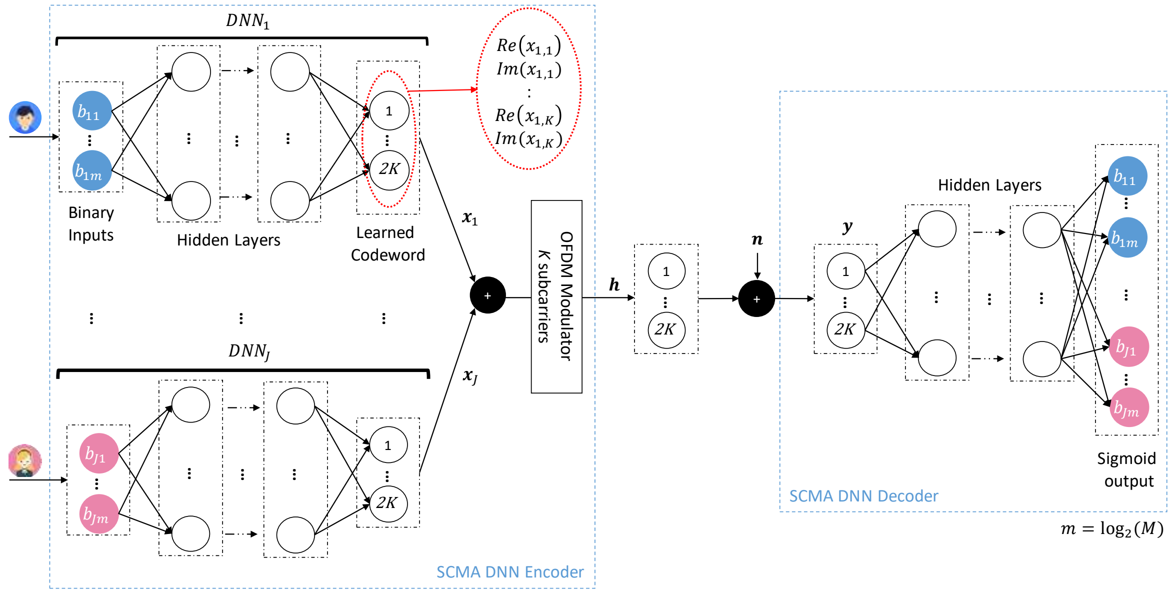

Figure 20. Structure of the SCMA autoencoder: a DNN with m nodes of input layer accepting binaries, and 2K nodes of output layer standing for the K-dimensional complex codeword, represent the SCMA mapping process for one single user. Multiple DNNs are stacked together to form the SCMA DNN encoder, another DNN is employed as a SCMA decoder.

order to propose a joint detection and decoding of channel-coded SCMA systems. The channel coding is randomly generated and is included in the training procedure which makes this proposition more suitable for practical applications.

A hybrid multiple access scheme was proposed in [173] where OFDMA orthogonal resources are used for near users and non-orthogonal SCMA based access is used for far ones. Two deep learning-based detectors were proposed, one DNN for near users detection, and the other is for far ones detection. In the two cases, a standard DNN structure was offline trained via a simulated data.

Data in NOMA based wireless systems will be more and more complex and heterogeneous which requires to employ powerful tools in order to inspect them and to detect the transmitted bits. This means that the deep learning based solutions will be increasingly attractive. However, there are still some challenges that must be overcome before the deep learning based NOMA systems become reality. Firstly, despite the fact that existing propositions enhance the NOMA performance but this come at the expense of high computational complexity which represents a major issue. Moreover, the optimal neural network, in terms of architecture, depth, number of parameters and training data and methods, is still a subject of research. Some perspectives and open directions of research will be discussed later in Section VII.

G. MIMO-SCMA Detectors

A straight-forward extension of SCMA factor graph to multiple-antenna case makes the number of resource nodes proportional to the number of antennas, this means that the complexity will respectively increase exponentially. Several research works try to overcome this difficulty [127], [129], [131], [133], [175]–[178]. Figure 21 represents a block diagram of an uplink MIMO-SCMA system when spatial multiplexing techniques is applied.

In [176] partial-decoding MPA was proposed, this method tries to remove redundant combinations at FNs to reduce decoding complexity, however, it still suffers from the exponential term in complexity. [131] introduced a stretched factor graph representation for MIMO-SCMA systems which enables the design of a hybrid BP and expectation propagation (EP) receiver that includes a channel decoder. Despite the relative low complexity of the stretch-BP-EP algorithm, its main shortcoming is that it can not be applied to user terminals which are equipped with multiple antennas. In [178], a minimum singular value estimation of the channel matrix is used to separate the resources into two categories according to the channel condition at each resource. Then, a low-complexity jointly Gaussian algorithm is applied to the well-conditioned resources, while the ML approach is used with the resources with bad-conditioned channel. This will allow a certain trade-off between performance and complexity when compared to the MPA detector. A similar approach based on an antenna-subcarrier subset selection was introduced in [127], the selection method is based on the channel norms. Furthermore, the MIMO-SCMA factor graph can be partitioned into subgraphs by a QR decomposition of the channel matrix such that the number of resource nodes depends only on the number of users sharing each resource, d_f, this can reduce the complexity of messages calculation [175]. Unlike the later research work, the authors in [133] associated the QR decomposition with EPA and consequently they proposed a sparse-channel-based EPA (SC-EPA), the SC-EPA intends to reduce the complexity by modifying the factor graph to create resource clusters, by exploiting the channel sparsity and by proposing a high-parallelism message passing technique. The effect of channel

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

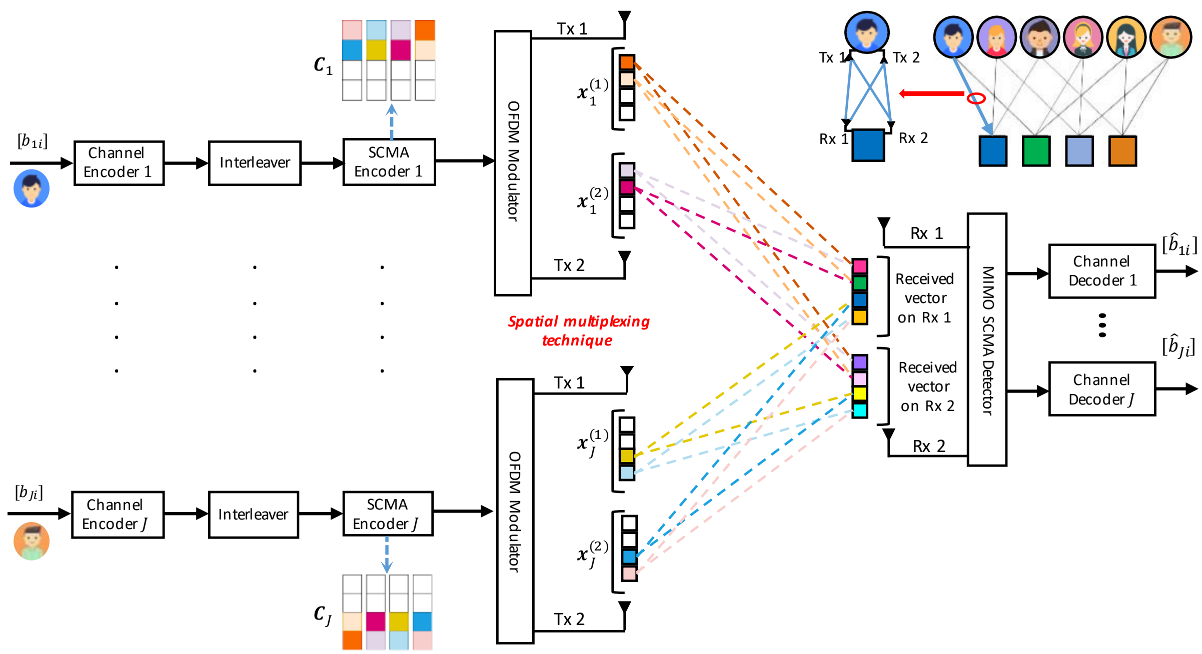

Figure 21. Block diagram of an uplink MIMO-SCMA system with $N_t = 2$ transmit antennas and $N_r = 2$ receive antennas.

estimation errors on the SC-EPA performance was studied, it seems as robust as other existing methods.

An outer-loop iterative structure based receiver for joint MIMO and multi-user detection was proposed in [177]. Authors introduced a basic-switching sliding window to separate users into active and silent ones during each iteration of the process such that only active user messages are exchanged during each inner-loop iteration of MPA. The Gaussian approximation was also used to further enhance the performance. This process guarantees the complexity reduction and the fairness among users since different users are activated for each outer-loop iteration. This method is called partially active message passing algorithm (PA-MPA). The former research work was extended in [129] where two GA-MPA methods were proposed to perform joint multi-user and MIMO detection. The first method updates the symbol likelihood ratios and it is denoted GA-MPA-S, while the second one, GA-MPA-B, is performed on the bit level. The Gaussian approximation results in some performance loss, the authors proposed to mitigate its effect by introducing a damping factor to compress the extrinsic information of each step of the outer-loop structure, namely the GA-MPA detector and the channel decoders. The complexity of proposed methods in [129] is linearly related to d_f and the number of antennas.

VI. Some other SCMA Aspects

In the above sections, we presented existing works on SCMA codebook and detector designs. Some other aspects will be covered in this section. We will first highlight existing radio and computing resources allocation strategies for SCMA systems. Afterwards, a review of SCMA implementation architectures, experiments and prototypes is introduced. Finally, the interplay between SCMA and other existing and emerging technologies will be investigated. The organisation of this section is illustrated in Figure 22.

Table IV
Key parameters in complexity analysis

Parameters	Description
N_t	Number of transmitter antennas
N_r	Number of receiver antennas
K	Spreading length (number of REs)
N	Codebook sparsity degree
N_{iter}	Total number of iterations
M	Codebook size
M_p	Number of projection points on the constellation
d_f	Degree of signal superposition on a given resource element
d_s	Maximum degree-of-freedom allowed in SIC-MPA receiver
J	Number of users
N_{OL}	Number of outer loops for MIMO-SCMA detection
N_{IL}	Number of inner loops for MIMO-SCMA detection
N_{v1}	Average number of visited layers for Type I (ISD method)
N_{v2}	Average number of visited layers for Type II (ISD method)
s	Sparsity level at the current iteration
PR	Pass rate (ratio of removed combinations to all combinations)

A. Resource Allocation

In literature, several works [179]–[181] was proposed to calculate the provision of users' data. Such as in [179], each RE is allocated to a cluster of users that share the same frequency resources in a NOMA-based system. Also, in [180], [181], authors considered associating IoT devices based on application requirements and resources availability. Each associated device is allocated to one RE with constant computing speed such that the computing resources of each RE were considered to be shared equally among all IoT devices.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

Table V
REVIEW OF EXISTING SCMA DETECTOR DESIGNS
DL = DOWNLINK, UL=UPLINK, SISO = SINGLE-INPUT SINGLE-OUTPUT, SIMO = SINGLE-INPUT MULTIPLE-OUTPUT, MISO = MULTIPLE-INPUT SINGLE-OUTPUT,
ONLY DOMINANT PART OF COMPLEXITY ORDER IS CONSIDERED, NOTATIONS ARE DEFINED IN TABLE IV,
$\surd \rightarrow$ ASSUMED TO BE KNOWN, $\times \rightarrow$ ASSUMED TO BE UNKNOWN.

Reference: Type	Reference: Authors	Year	Approach	Scenario	CSI	Add & Mult Complexity	Description
MPA based detectors	Zou *et al.* [112]	2015	SIC-MPA	SISO DL	$\surd$	Neglected & $\mathcal{O}(N_{\text{iter}} N_r K M_p^{d_s})$	Joint use of SIC and MPA. Only active users per RE are considered. Specific constellation design is needed to be employed.
	Zhang *et al.* [140], [141]	2016 2018	DMPA	SISO	$\surd$	$\mathcal{O}(d_f^3 ln(d_f))$ & $\mathcal{O}(d_f^3 ln(d_f))$	Requires a small d_f which limits the choice of codebook. Based on PDF discretizing and FFT in VNs.
	Yang *et al.* [116]	2016	Threshold-based MPA	SISO UL	$\surd$	$\mathcal{O}(N_{\text{iter}} JNM^{d_f-1} d_f)$ & $\mathcal{O}(2N_{\text{iter}} JNM^{d_f} d_f)$	Users are divided into two groups based on the codeword reliability which is integrated in the iterative process. Inconvenient : loss of posterior information of users using the same RE.
	Qi *et al.* [108]	2017	Max-log-MPA	SISO UL	$\surd$	$\mathcal{O}(2KM^{d_f} d_f))$ & 0	Approximating the logarithm of a sum of exponential operations into a maximum operation.
	Tian *et al.* [144]	2017	Improved Log-MPA	SISO DL	$\surd$	$\mathcal{O}(2JM(M^{d_f-1}-1))$ & 0	Log-MPA adds a correction term to MAX-log-MPA. The authors improved Log-MPA by updating the FNs only within a restricted search region.
	Wu *et al.* [176]	2017	PD-MPA	MIMO UL	$\surd$	$\mathcal{O}(NN_{\text{iter}}(M-1)M^{d_f-1} d_f^2 .PR)$ & Neglected	Removing redundant combinations while updating the VNs. Based on Log-MPA.
	Huang *et al.* [126]	2017	GenA-MPA	MIMO UL	$\surd$	Neglected & $\mathcal{O}(JK^2 + KN_{\text{iter}} M_p^{d_f})$	Approximated priori probability transforms the vector estimation problem into a scalar one, it is suitable for hardware implementation.
	Huang *et al.* [126]	2017	SIC-GenA-MPA	MIMO UL	$\surd$	Neglected & $\mathcal{O}(JK^2 + N_{\text{iter}} M_p^{d_f})$	Extends GenA-MPA by eliminating the calculated user layer, the complexity is reduced.
	Dai *et al.* [111]	2017	SS-MPA MS-MPA	SISO UL	$\surd$	Not specified & Not specified	A lookup table is introduced to substitute the information calculation. SS-MPA jointly updates the message at FNs and VNs. MS-MPA executes different updates in parallel.
	Lai *et al.* [155]	2018	LDPC-MPA	SISO UL	$\surd$	Significant complexity reduction	Joint factor graph for SCMA and LDPC channel decoding. MPA was simplified by using partial message passing.
	Gao *et al.* [113]	2018	CS-MPA	SISO	$\surd$	Not specified & $8K + 3KN + 2Ns^2 + s^3$	The detector is divided into two phases : MPA with few iterations followed by a sparse error correction. The main complexity lies in MPA.
	Han *et al.* [114]	2018	Serial MPA	SISO	$\surd$	Not specified & Not specified	Updates the probability of user output codeword immediately when the information of one of its corresponding FNs is updated. EXIT charts are used to analyze MPA performance.
	Dai *et al.* [177]	2018	PA-MPA	MIMO UL	$\times$	Neglected & $\mathcal{O}(N_{\text{OL}} N_{\text{IL}} \sum_{n=1}^{N_r} 2^J)$	An outer-loop iterative process for joint MIMO and multi-user detection. A basic-switching sliding window to separate users into active and silent ones during each iteration.
	Jia *et al.* [121]	2018	PM-MPA	SISO UL	Not speci-fied	$\mathcal{O}(M^{d_f} NJ(N_{\text{iter}} - t))$ & $\mathcal{O}((N_{\text{iter}}(d_f-1)M^{d_f} NJ)$	The codewords of a given number of chosen users are determined after a predefined number of iterations. Only the message of the undetermined users are updated in the remaining iterations.
	Lai *et al.* [122]	2018	EML-MPA	SISO UL	$\surd$	Neglected & $\mathcal{O}(M_p^{d_f})$	Trellis representation is introduced by mapping SCMA constellation to a Galois field. A truncated-messages based detector is based on selecting the most reliable messages.
	Ye *et al.* [96]	2019	Q-MPA	SISO UL	$\surd$	The cost function is calculated $\mathcal{O}\left(\sqrt{M_p^{d_f+1}}\right)$ times	Algorithm-aided MPA is accelerated by exploiting the results of a quantum search which is applied at the FNs.
	Ma *et al.* [115]	2019	Edgewise serial MPA	SISO UL	$\surd$	$\mathcal{O}(Kd_f[(d_f+1)M^{d_f} - M])$ & $\mathcal{O}(Kd_f M^{d_f}(2d_f+1))$	This method updates messages in an edge by edge way such as a more reliable extrinsic information is used in a balanced manner.
	Shi *et al.* [124]	2019	Log-MPA	SISO UL	$\surd$	Not specified & Not specified	MPA with flexible number of iterations. The iterative process continues until the codeword convergence rate is higher than a threshold. A balance between BER and complexity can be achieved.
	Dai *et al.* [129]	2019	GA-MPA-S GA-MPA-B	MIMO UL	$\times$	Neglected & $\mathcal{O}\left(2N_{\text{iter}} d_f M\right)$	Gaussian-approximation based MPA. GA-MPA-S updates the symbol likelihood ratios and GA-MPA-B is performed on bit level. Complexity was pulled down from exponential to linear order.
	Peng *et al.* [123]	2020	EML-MPA	SISO UL	$\surd$	$\mathcal{O}(KM^{d_f} d_f)$ & $\mathcal{O}(KM^{d_f}(d_f+2))$	Dynamic trellis based message passing algorithm is proposed such as the truncation is dynamically decreased as the iterations progress. A channel adaptive version of this algorithm is also introduced.

Reference		Year	Approach	Scenario	CSI	Add & Mult Complexity	Description
Type	Authors						
EPA based detectors	Meng *et al.* [130]	2017	EPA	SIMO UL	✓	$\mathcal{O}(N_{\text{iter}} N_r K M d_f)$	This method is based on approximate Bayesian interference. Instead of continuous exponential terms, message passing is reduced to mean and variance calculation of approximate Gaussian distributions.
	Yuan *et al.* [131]	2018	Stretch-BP-EP	MISO DL	×	$\mathcal{O}(N_{\text{iter}}(N_t + J)$	Stretched factor graph representation that enables the design of a hybrid BP and EP receiver with channel decoder. It can be only applied to user terminals with single antenna.
	Chen *et al.* [132]	2018	Bayesian Interference	SISO UL	✓	$\mathcal{O}(N(d_f + M))$ & $\mathcal{O}(MNd_f)$	The authors proposed a Monte Carlo Markov chain based SCMA detection algorithm. A new joint update parallel probability sampler is proposed.
	Wang *et al.* [133]	2020	SC-EPA	MIMO UL	✓	Neglected & $\mathcal{O}(Nd_f N_{\text{iter}}(6N_r + 5M_p))$	QR decomposition is combined with EPA which allows to exploit the channels sparsity. A high-parallelism message passing technique is proposed. The effect of channel estimation errors is studied.
SD based detectors	Yang *et al.* [145]	2017	SD	SISO UL	✓	$\mathcal{O}(d_f K M^{d_f})$ & $\mathcal{O}(KM^{d_f}(d_f + 1))$	Space restriction of codeword research to a hyper-sphere centered around the received signal whose radius is adjusted according to the noise power at the receiver. Method with high complexity in the low SNR region.
	Wei *et al.* [81], [143]	2016 2017	LSD	SISO UL	✓	$\mathcal{O}(2N(d_f \log_2(M))^3)$ & Neglected	The search is limited to a list of all the candidate lattice points within a given radius around the received vector.
	Vameghestahbanati *et al.* [146]	2017	MSD	SISO UL	✓	$\mathcal{O}((4d_f + 2)N_{v1} + 2N_{v2})$ & $\mathcal{O}((4d_f + 2)N_{v1} + 2N_{v2})$	Based on a modified tree search method. Using the Tikhonov regularization to facilitate applying SD to SCMA systems.
	Li *et al.* [150]	2018	MSTS	SISO UL	✓	Lesser number of cost function calculations	Modified single tree search avoids the unnecessary node searching by sorting channel matrix in ascending order and by employing a non-zero low bound enumeration.
	Ye *et al.* [96]	2019	QSD-MPA	SISO UL	✓	The cost function is calculated less than $\mathcal{O}\left(\sqrt{sM_p^{d_f+1}}\right)$ times	A quantum search is used to identify all possible codeword combinations within a given hyper-sphere, thereafter MPA is applied on identified points only.
	Vameghestahbanati *et al.* [151]	2019	ISD	SISO UL	✓	$\mathcal{O}((4d_f + 2)N_{v1} + 2N_{v2})$ & $\mathcal{O}((4d_f + 2)N_{v1} + 2N_{v2})$	This method extends the MSD one to support any arbitrary regular or irregular constellation topology.
Other SCMA detectors	Sun *et al.* [156]	2019	LDPC-coded SCMA	SISO UL	✓	$\mathcal{O}(660 N_{\text{IL}} N_{\text{OL}})$	The interface between the SCMA detector and the LDPC decoder is optimized such as a proof of concept was implemented using 40nm CMOS technology.
	Jiao *et al.* [159]	2020	Polar-coded SCMA	SISO UL	×	$\mathcal{O}(N_{\text{iter}} J d_f M^{d_f})$	Iterative joint channel estimation and decoding scheme is proposed where Max-Log-MPA and S-SCL algorithms are used, respectively, for SCMA and polar decoding.
	Lu *et al.* [171]	2018	Deep learning	SISO UL	×	Not specified	A DNN was designed such as the propagation between two of its layers is calculated based on how messages are passed between FNs and VNs in traditional iterative MPA.
	Lin *et al.* [174]	2020	Deep learning	SISO UL	×	Not specified	Autoencoders are employed to automatically design SCMA codebooks and construct the corresponding detetor through AWGN channels.

It is important to remind that one of reasons behind using SCMA to accommodate the massive connectivity for IoT systems, comes from the scarcity of frequency resources. Moreover, SCMA can improve network connectivity and maximize data rate provision [182], [183]. In other studies [70], [71], [184], [185], SCMA parameters were investigated both to reach the high quality of experience (QoE) requirements of IoT devices, and to explore the various performance aspects such as connectivity, throughput, task completion time, and complexity. Therefore, it is important to fit with the specific needs of each particular device to maintain satisfactory QoE in terms of computing and radio resources allocation. Thus, adaptive SCMA schemes with flexible resource scheduling were proposed in [72], [73], [184].

The problem of data rate provision, based on SCMA resource management, while still satisfying the delay requirements of IoT devices is formulated as data rate maximization problem in [185]–[190]. The appropriate solution was to divide the non-convex problem into three subproblems: 1) power allocation subproblem, where each BS undertakes power optimization of associated users, 2) codebook assignment subproblem, in which every user is associated to a codebook that provides the highest data rate considering the combined effect of all subcarriers within that codebook, and 3) joint power allocation and codebook assignment problem. In the following we highlight these subproblems with their suboptimal proposed solutions.

Power Allocation Subproblem: A majority among SCMA studies adopted the uniform power distribution scheme, which is the simplest method of power allocation. Assuming that the

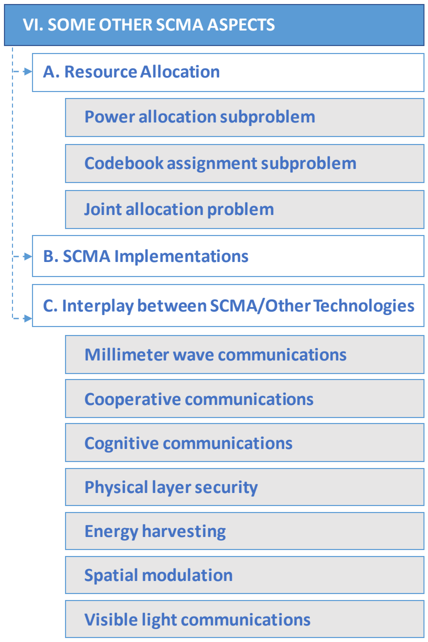

Figure 22. The organization and structure of Section VI.

total transmit power is P^{tot} and the power allocated to each subcarrier k is

$$P_k = \frac{P^{tot}}{K}, \tag{8}$$

where K is the number of subcarriers. The ergodic system capacity can be calculated as

$$C = K\mathbb{E}\left\{ \log_2 \left(1 + \frac{h_k \cdot \frac{P^{tot}}{K}}{N_0} \right) \right\} \tag{9}$$

where h_k denotes the channel gain of the associated subcarrier k. The fact that the power is equally distributed means that the channel state is not taken into account, this is not optimal.

However, when the CSI is known at the receiver, the water filling algorithm [191] seems to be an appropriate solution in terms of system capacity. The idea is that when CSI differs across users of an orthogonal multiple access based communication system, the system capacity can be maximized based on the water filling algorithm under the assumption that the transmission power is limited. Thus, with perfect knowledge of the users' channel states equation (8) becomes,

$$P_k = \max\left(\gamma - \frac{h_k}{N_0}, 0\right), 1 \leq k \leq K, \tag{10}$$

where γ is a constant that is determined by the power constraint condition and that makes the total transmission power equal to $\sum_{k=1}^{K} P_k = P^{tot}$. The ergodic system capacity is then given as,

$$C = \mathbb{E}\left\{ \sum_{k=1}^{K} \log_2 \left(1 + \frac{h_k P_k}{N_0} \right) \right\} \tag{11}$$

Nevertheless, the water filling algorithm still can not be applicable to NOMA systems since it can not guarantee the system fairness and it can even prevent some users to communicate with the BS. Motivated by these issues, a power allocation scheme that ensure users' fairness for multi-user SCMA downlink systems is proposed in [186] using a low-complexity polynomial algorithm in order to enhance the performance compared to the equal power allocation scheme. Later, a maximum capacity power distribution scheme is proposed in [190] in order to maximize the system capacity and to satisfy users' minimum rate requirement through three-level power allocation model consisting of single-user subcarrier, intragroup and intergroup power allocations.

Codebook Allocation Subproblem: Here, we want to explain how codebook allocation was introduced in some of existing works [187]–[189]. In [187], the REs were assigned to users based on their CSI, this process is considered as codebook assignment since it will define the factor graph matrix. A comparative study among three codebook assignment methods was proposed. The first one is fixed user order (FUO) algorithm where users are sorted in a fixed arbitrary order. A group of N subcarriers with larger gains is selected among all available subcarriers and assigned to the first user. Then the procedure is repeated for the next one until all users are assigned to their corresponding subcarriers. The serial scheme of this method limits the system performance. To tackle this issue, instead of sorting users arbitrarily, an opportunistic assignment (OA) algorithm is proposed. That is the first user is set as the one with the best overall quality of subcarriers' gains based on a specific metric such as the root mean-square of subcarrier gains for instance. The rest of the process is similar to FUO. Despite of its better performance compared to FUO, OA method may not provide an acceptable fairness in terms of individual user's BER even when providing the best overall average BER. Motivated by this issue, proportional fairness (PF) assignment method can be an appropriate solution. Each user's current CSI as well as a history of several past CSI values are employed as performance and fairness indicators in order to provide the necessary trade-off based on a priority function such that the user has the priority only when the past CSI values have not been ranked highest among all users.

Other resource allocation strategy for downlink SCMA was studied in [188], namely largest weighted delay first method which represents an improved version of classical PF one. This method is associated to a priority function for each subcarrier which takes into consideration time delay of data packets, traffic priority and CSI of each user. Furthermore in [189], the SCMA resource allocation process is decomposed into two steps: codebook combination and codebook allocation. Firstly REs are combined into as many high-quality codebooks as possible based on users' CSI. Then, user equipment's are assigned to the combined codebooks based on coalitional game which aim is to maximize the intra-group fairness.

Joint Power Allocation and Codebook Assignment Subproblem: In addition to various propositions to enhance resource management using power allocation or codebook assignment schemes, several research works, such as in [185], [192]–[194], proposed a joint power and codebook assignment.

In [194], an iterative jointly codebook and power assignment algorithm, which prioritizes the energy efficiency, was formulated. The algorithm was divided into two subproblems. Firstly, codebooks are assigned under a given power alloca-

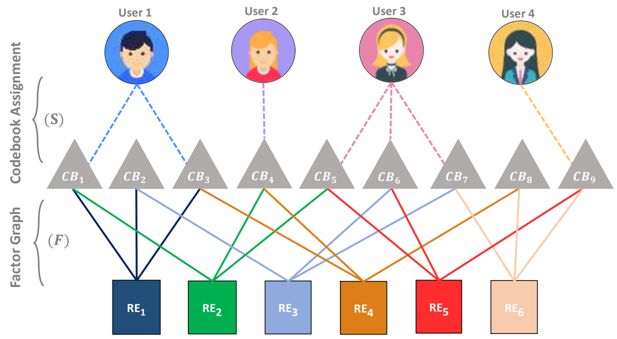

Figure 23. An example of mapping relationship among users, codebooks, and subcarriers in [72]: the number of REs is 6, the number of available codebooks is 9, the number of active users is 4, and $d_f = 3$.

tion and then the power is allocated under given codebook assignment. The energy efficiency is maximized by alternately iterating through the two phases. Also, a SCMA scheme for edge computing was proposed in [185]. The aim was to maximize the data rates, under the maximum power constraint of the BS, using SCMA based heterogeneous networks. This objective is reached by solving two allocation problems. A power allocation problem, which can be solved using the water filling technique, and a codebook allocation problem resolved by assigning codebooks with the highest signal to interference plus noise ratio to SCMA users on a first-come-first-serve basis. Furthermore, subcarrier reuse constraints was also investigated in [192].

Other codebook assignment strategies were proposed with the aim of assigning multiple codebooks for each user [72], [195], [196] or reusing the same codebook by different users [197]. For instance, the research work in [72] focus on the energy saving problem in the uplink and downlink SCMA systems based on a cross-layer approach which takes into account the user features, i.e. quality of service requirement and channel state. The authors introduced the codebook assignment matrix, $\mathbf{S}$, where $[\mathbf{S}]_{i,j} = 1$ if codebook i is assigned to user j otherwise $[\mathbf{S}]_{i,j} = 0$ as shown in Figure 23. The idea is to try to optimize, at the same time, the factor graph matrix, the codebook assignment matrix and how power is allocated. The formulated problem is an integer linear program that is NP-hard. It was solved in a direct way using dual coordinate search, i.e. by using weighting coefficients that can coordinate the codebook assignment among users, this is considered as a Lagrangian relaxation of the original problem. This method avoids the relaxed linear program as an intermediate step which reduces the computational complexity significantly.

B. SCMA Implementations

Several hardware implementation strategies of SCMA detectors exist in the state-of-the-art [108], [120], [156], [198]–[200]. A fixed-point implementation of the Log-MPA was studied in [198], the results were verified via an Altera Stratix V FPGA evaluation board. A thread-based parallelism methodology was introduced in [108] to implement the iterative Max-Log-MPA, a simulation via a NVIDIA TITAN X (Pascal) and CUDA was performed. A stochastic computing based SCMA detector is proposed in [200]. The idea is to simplify the complex computations by representing the numbers by random bit streams. The authors proposed a new bit stream generator before designing stochastic FN update and stochastic VN update. The designed detector was also synthesized with SIMC 65 nm CMOS technology. In [120], [199], a stage-level scheduling optimization and folding techniques were proposed in order to conceive a hardware architecture of the Max-Log deterministic MPA. The proposed detector was implemented on Xilinx Virtex-7 XC7VX690 T FPGA. The first application-specific integrated circuit which supports multi-user iterative detection and decoding for LDPC-coded SCMA systems was introduced in [156]. A proof of concept was implemented using 40nm CMOS technology, a throughput of 1.198 Gb/s and 599 Mb/s for 4×6 and 8×12 SCMA systems was, respectively, achieved.

C. Interplay between SCMA and Other Technologies

SCMA can be flexibly combined with many existing wireless technologies and emerging ones including massive MIMO, mmWave communications, cognitive and cooperative communications, visible light communications, physical layer security, energy harvesting, spatial modulation, and so on. In this subsection, we present the interplay between SCMA and the above technologies which can further increase scalability, spectral efficiency, energy efficiency, and greenness of future communication networks.

- *Millimeter wave communications:* Due to the high data rate requirements of future wireless networks, the communication at mmWave bands (30 GHz to 300 GHz) has been subject to intensive research during the past decade [201]–[205]. However, since mmWave channels are sparse in spatial domain, the number of simultaneous connections at these very high frequencies has shown to be limited. Motivated by this issue, Thang *et al.* highlighted the coexistence of SCMA and mmWave in [204], [205] where SCMA was introduced with beam-space mmWave MIMO systems for high SNR regions. However, it is worth noting that the number of supported users cannot be larger than the number of radio frequency chains in hybrid MIMO architecture.
- *Cooperative communications:* The main idea of cooperative communications is to share well the resources among different nodes of the wireless network in order to improve its total throughput. With different relaying schemes, device-to-device communication, and multi-cell cooperative transmission, SCMA and cooperative communications are merged together [206]–[208]. A novel two-tier downlink cooperation system with limited requirements on synchronization was proposed in [206] where SCMA scheme is seen as a suitable technique for the lower tier of multi-cell transmission to achieve a multi-site distributed cooperation. Moreover, relay-aided SCMA recently attracted considerable attention for its great promise in improving spectral efficiency and user fairness. An optimization of resource allocation in dual-hop relay-assisted uplink SCMA network was proposed

in [207] where user equipment transmits information to a relay station via SCMA codebook assignment and then the relay station amplifies-and-forwards the information to the base station through OFDMA subcarrier pairing.

- *Cognitive communications:* Cognitive radio promises a great potential for spectral efficiency and has been under investigation for more than two decades. However, SCMA was applied to cognitive radio networks to increase the number of users and further increase the spectral efficiency [203], [209], [210]. In [203], cognitive radio capabilities were enabled for SCMA systems operating in mmWave bands by sensing the spectrum holes and adapting the transmission. In [210], a centralized cooperative spectrum sensing scheme is employed to detect the primary users in SCMA systems. Log-MPA detector calculates the soft information which is then passed to a fusion center where the Wald-hypothesis test is employed to decide on the presence of a primary user.
- *Physical layer security:* Physical layer security techniques exploit physical characteristics of wireless communication channel, for instance noise, fading, and interference, to guarantee secure communications directly at the physical layer [211]. However, a little research is conducted in the case of large number of users, and limited power and frequency resources. That is why [212] proposes a physical security scheme for SCMA-based systems in order to optimize power allocation, security capacity and number of users. In [213], employing randomized codebook generation method based on random angles extracted from channel phases can ensure the security of downlink SCMA but requires the knowledge of CSI.
- *Energy harvesting:* Energy harvesting communications is an essential enabler of green self-sufficient and self-sustaining communications [214]. In order to control harvested energy from ambient radio frequency electromagnetic signals, a combination of SCMA and energy harvesting communication is proposed in [193], [215] where a trade-off between data rate and energy consumption of SCMA networks is investigated when wireless power transfer is employed. The authors formulated a weighted rate and energy minimization optimization problem while jointly considering power allocation, codebook assignment and power splitting.
- *Spatial modulation:* Spatial modulation provides a trade-off between spectral and energy efficiency. It is based on a simple appealing design which consists in *implicitly* transmitting some additional bits through the indices of active antennas [216]. Spatial modulation aided by SCMA was recently proposed in [217]–[220]. The principle is that the *explicit* bits of each user are mapped directly to a codeword drawn from a specific codebook [218]. The authors in [217] derived the best a posteriori detector based on which they develop a reduced complexity MPA-aided detector. The previous works require an integer power of two transmit antennas, this drawback was rectified in [219] by introducing a rotational generalized spatial modulation-SCMA system, the authors' work was extended in [220] where three low-complexity detection algorithms were proposed.
- *Visible light communications:* The visible light communications are suitable for solving spectrum congestion and scarcity issues by shifting transmission frequencies from conventional radio frequency ranges to the visible light range [221]. To provide an efficient way to handle the available resources in multi-user scenarios, the interplay between SCMA and visible light communications has been recently studied [222]–[227]. For instance, in [226], [227], SCMA codewords are superimposed in the color domain where transmit antennas become light emitting diodes and receive antennas become photodetectors, this will minimize error propagation when compared to PD-NOMA.

VII. Discussion and Open Directions of Research

Existing SCMA schemes are capable of improving the spectral efficiency of wireless systems based on non-orthogonal resources sharing. SCMA mechanisms are designed to facilitate the support of massive connectivity and are considered as potentially promising multiple-access candidate for future generations. However, there are still numerous challenging problems to be solved. Hence below some of the key challenges of SCMA will be identified, along with opportunities and future research trends addressing these challenges.

- *MIMO-SCMA*: Few existing works studied MIMO-SCMA systems [129], [131], [133], [177], [178], however it will be useful to extend these works to further examine the MIMO-SCMA detection problem by applying techniques like sphere decoding or advanced search algorithms for instance. Furthermore, some extensive research is needed to investigate the interplay between SCMA and some other MIMO aspects, for instance space-time block coding, beamforming, massive number of antennas, etc.
- *Joint all-in-one signal detector*: The joint channel-coded SCMA detectors and the joint MIMO-SCMA ones were reviewed in Subsections V-E and V-G respectively. It is promising to extend this effort to conceive joint all-in-one signal detector. That is to propose a joint factor graph of MIMO, polar/LPDC decoder and SCMA detector, and all-in-one algorithm to exploit this factor graph. To the best of our knowledge, this matter is not yet studied.
- *Deep Learning*: One solution for SCMA codebook design and/or signal detection is to exploit machine learning techniques and more specifically deep learning which proposes several efficient tools. For the time being, this axe of research was rarely studied [169]–[173], that is why it is deemed as promising and attractive. Some deep architectures such as deep convolutional neural networks, recurrent neural networks and Bayesian neural networks are yet to be associated to SCMA, this requires some investigations. In addition, deep reinforcement learning is never used in the SCMA case, this mechanism can optimize the performance of SCMA detectors. Furthermore, existing works employ simulated data for training which is not ideal, therefore it is important to collect a large

amount of real SCMA data in order to build a training database which allows a more powerful offline training. However, the offline training can be aided by an online training which is based on transmitting some pilot signals in real-time, this is a significant research issue for the future.

- *Channel estimation*: Most of the aforementioned SCMA detection methods assume the perfect knowledge of CSI which is not realistic and will lead to performance degradation, this problem can be overcome either by studying the effect of CSI estimation errors [133], or by carrying out a joint channel estimation and data decoding methods [136], [159]. The impact of CSI estimation errors on the performance of existing algorithms must be further evaluated such that a fair comparison among them will be feasible, and new improved ones must be proposed if necessary.
- *Synchronisation*: Existing SCMA detectors assume a perfect synchronisation at the receiver which it is not realistic. This can lead to some performance degradation especially in the SCMA uplink systems [135]. For instance, a delayed signal from a user will disturb the message exchange in MPA based detectors. An additional effort must be dedicated to proposing a more robust detectors which can cope with synchronisation problems.
- *Adaptive SCMA*: In regular SCMA, the users are uniformly served which it is not optimal since it will be more practical if the SCMA systems can be differently adapted to users needs and their channel states for instance. Exiting works on adaptive or irregular SCMA are reviewed in III-B, however this area must be further explored to propose some tailored codebook designs and especially tailored detectors for adaptive SCMA systems.
- *Peak-to-average power ratio*: Few papers studied the problem of high peak-to-average power ratio of SCMA systems which leads to signal distortion. One solution is proposed in [228] which consists in applying a clipping technique. However this produces an extra noise which will be added to the AWGN, the clipping noise characteristics were analysed such that the MPA detector can be modified to integrate the two noise types and an extra performance degradation can be avoided. This problem must be further investigated.
- *Implementations*: A quick glance on existing SCMA implementations shows that the case of MIMO-SCMA systems was not studied which will be required to hope that SCMA can be integrated in the future wireless telecommunication systems.
- *Reconfigurable intelligent surface:* The emerging technology of reconfigurable intelligent surface (RIS) for wireless communications can be employed to effectively control the radio propagation environment by adjusting the signal's waveform parameters, for instance, phase and amplitude, based on configuring some low-cost reflecting elements [229]. The applications of the RIS technique in PD-NOMA systems was recently introduced in [230], [231]. On the other hand, its application in CD-NOMA systems is not yet examined except to one research work in which the authors proposed RIS-assisted SCMA system [232]. The interplay between RIS and NOMA, especially SCMA, must be further investigated.
- *Hybrid multiple access:* Subsection II-B listed existing SCMA based hybrid multiple access schemes. Performance analyses of the proposed hybrid schemes are satisfactory such that these schemes can be considered as suitable candidates for next generations of wireless systems. However, this will not be possible if some additional work is not done to design robust low-complexity joint detectors which take into consideration the realistic impact of channel state. Furthermore, one interesting perspective is to study users clustering into groups such that orthogonal codes are employed for inter-group separation while SCMA codebooks are used inside each group.

VIII. Conclusions

This survey hopes to be a useful review of the state-of-the-art of SCMA techniques. SCMA encoder and detector designs were first comprehensively presented. Afterwards, we provided essential elements on some other SCMA aspects as resource allocation, its interplay with some other technologies, and existing prototypes. However, for deeper understanding of the details of each point in this survey, the reader is recommended to further study the original papers and documents.

Appendix

To better illustrates SCMA mapping, a numerical example of complete SCMA codebooks, as described in [91] and depicted in Figure 11 and Figure 13, is provided in the following.

$$\mathbf{CB}_1 = \left(\begin{bmatrix} 0 \\ -0.182-0.132i \\ 0 \\ +0.785 \end{bmatrix} \begin{bmatrix} 0 \\ -0.635-0.462i \\ 0 \\ -0.224 \end{bmatrix} \begin{bmatrix} 0 \\ +0.635+0.462i \\ 0 \\ +0.224 \end{bmatrix} \begin{bmatrix} 0 \\ +0.182+0.132i \\ 0 \\ -0.785 \end{bmatrix} \right)$$

$$\mathbf{CB}_2 = \left(\begin{bmatrix} +0.785 \\ 0 \\ -0.182-0.132i \\ 0 \end{bmatrix} \begin{bmatrix} -0.224 \\ 0 \\ -0.635-0.462i \\ 0 \end{bmatrix} \begin{bmatrix} +0.224 \\ 0 \\ +0.635+0.462i \\ 0 \end{bmatrix} \begin{bmatrix} -0.785 \\ 0 \\ +0.182+0.132i \\ 0 \end{bmatrix} \right)$$

$$\mathbf{CB}_3 = \left(\begin{bmatrix} -0.635+0.462i \\ +0.139-0.176i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} +0.182-0.132i \\ +0.487-0.616i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -0.182+0.132i \\ -0.487+0.616i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} +0.635-0.462i \\ -0.139+0.176i \\ 0 \\ 0 \end{bmatrix} \right)$$

$$\mathbf{CB}_4 = \left(\begin{bmatrix} 0 \\ 0 \\ +0.785 \\ -0.006-0.224i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -0.224 \\ -0.019-0.785i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ +0.224 \\ +0.019+0.785i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -0.785 \\ +0.006+0.224i \end{bmatrix} \right)$$

$$\mathbf{CB}_5 = \left(\begin{bmatrix} -0.006-0.224i \\ 0 \\ 0 \\ -0.635+0.462i \end{bmatrix} \begin{bmatrix} -0.019-0.785i \\ 0 \\ 0 \\ +0.182-0.132i \end{bmatrix} \begin{bmatrix} +0.019+0.785i \\ 0 \\ 0 \\ -0.182+0.132i \end{bmatrix} \begin{bmatrix} +0.006+0.224i \\ 0 \\ 0 \\ +0.635-0.462i \end{bmatrix} \right)$$

$$\mathbf{CB}_6 = \left(\begin{bmatrix} 0 \\ +0.785 \\ +0.139-0.176i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ -0.224 \\ +0.487-0.616i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ +0.224 \\ -0.487+0.616i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ -0.785 \\ -0.139+0.176i \\ 0 \end{bmatrix} \right)$$

The 2-dimensional codebooks with 4-codewords, generated for $J = 6$ users, as described in [91].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

REFERENCES

[1] K. Routh and T. Pal, "A survey on technological, business and societal aspects of internet of things by q3, 2017," in *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*, 2018, pp. 1–4.

[2] L. Dong, H. Zhao, Y. Chen, D. Chen, T. Wang, L. Lu, B. Zhang, L. Hu, L. Gu, B. Li, H. Yang, H. Shen, T. Tian, Z. Luo, and K. Wei, "Introduction on imt-2020 5g trials in china," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 8, pp. 1849–1866, Aug. 2017.

[3] J.Rodriguez, *Fundamentals of 5G Mobile Networks*. Wiley, 2014, ch. 10, pp. 221–240.

[4] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[5] N. Abu-Ali, A. M. Taha, M. Salah, and H. Hassanein, "Uplink scheduling in lte and lte-advanced: Tutorial, survey and evaluation framework," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1239–1265, 2014.

[6] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5g and beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec 2017.

[7] Z. Wei, J. Yuan, D. W. K. Ng, M. Elkashlan, and Z. Ding, "A survey of downlink non-orthogonal multiple access for 5g wireless communication networks," *ZTE Communications*, no. 4, pp. 17–26, 2016.

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

[9] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (noma) systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[10] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink noma systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7244–7257, 2016.

[11] Y. Wu, C. Wang, Y. Chen, and A. Bayesteh, "Sparse code multiple access for 5g radio transmission," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–6.

[12] A. Ghaffari, M. Leonardon, A. Cassagne, C. Leroux, and Y. Savaria, "Toward High-Performance Implementation of 5G SCMA Algorithms," *IEEE Access*, vol. 7, pp. 10 402–10 414, Jan 2019.

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

[14] M. Vameghestahbanati, I. D. Marsland, R. H. Gohary, and H. Yanikomeroglu, "Multidimensional constellations for uplink scma systems—a comparative study," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2169–2194, thirdquarter 2019.

[15] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5g," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2294–2323, JUL-SEP 2018.

[16] M. Aldababsa, M. Toka, S. Gokceli, G. K. Kurt, and O. Kucur, "A tutorial on nonorthogonal multiple access for 5g and beyond," *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–24, Article ID 9713450.

[17] M. Basharat, W. Ejaz, M. Naeem, A. M. Khattak, and A. Anpalagan, "A survey and taxonomy on nonorthogonal multiple-access schemes for 5g networks," *Transactions on Emerging Telecommunications Technologies*, vol. 29, no. 1, p. e3202, 2018.

[18] Z.-g. Ding, M. Xu, Y. Chen, M.-g. Peng, and H. V. Poor, "Embracing non-orthogonal multiple access in future wireless networks," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 3, pp. 322–339, 2018.

[19] N. Ye, H. Han, L. Zhao, and A.-h. Wang, "Uplink nonorthogonal multiple access technologies toward 5g: A survey," *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–26, Article ID 6187580.

[20] Y. Cai, Z. Qin, F. Cui, G. Y. Li, and J. A. McCann, "Modulation and multiple access for 5g networks," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 629–646, 2018.

[21] M. Vaezi, G. A. Aruma Baduge, Y. Liu, A. Arafa, F. Fang, and Z. Ding, "Interplay between noma and other emerging technologies: A survey," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 4, pp. 900–919, 2019.

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

[23] A. Anwar, B.-C. Seet, M. A. Hasan, and X. J. Li, "A survey on application of non-orthogonal multiple access to different wireless networks," *Electronics*, vol. 8, no. 11, p. 1355, Nov 2019.

[24] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain noma with enabling technologies of future wireless networks," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2192–2235, 2020.

[25] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive non-orthogonal multiple access for cellular iot: Potentials and limitations," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 55–61, 2017.

[26] A. Benjebbour, "An overview of non-orthogonal multiple access," *ZTE Communications*, vol. 15, no. S1, p. 21, 2017.

[27] Y. Shan, C. Peng, L. Lin, Z. Jianchi, and S. Xiaoming, "Uplink multiple access schemes for 5g: A survey," *ZTE Communications*, vol. 15, no. S1, p. 21, 2017.

[28] Z. Ma and J. Bao, "Sparse code multiple access (scma)," in *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Springer, 2019, ch. 12, pp. 369–416.

[29] B. Makki, K. Chitti, A. Behravan, and M. S. Alouini, "A Survey of NOMA: Current Status and Open Research Challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.

[30] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh, "Machine-type communications: current status and future perspectives toward 5g systems," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 10–17, 2015.

[31] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, 2013, pp. 770–774.

[32] D. Cai, Z. Ding, P. Fan, and Z. Yang, "On the Performance of NOMA With Hybrid ARQ," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 10 033–10 038, 2018.

[33] B. Narottama and S. Y. Shin, "Dynamic Power Allocation for Non-Orthogonal Multiple Access with User Mobility," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2019, pp. 0442–0446.

[34] B. Makki, M. Hashemi, and A. Behravan, " Smart HARQ in NOMA-Based Networks," document PCT/EP2018/062393, Ericsson, Stockholm, Sweden, May 2018.

[35] ——, "CSI-Constrained User Pairing in Dense Uplink NOMA Networks," document PCT/EP2018/053733, Ericsson, Stockholm, Sweden, Feb. 2018.

[36] C. T. Rodenbeck, M. Martinez, J. B. Beun, J. Silva-Martinez, A. I. Karsilayan, and R. Liechty, "When Less Is More . . . Few Bit ADCs in RF Systems," *IEEE Access*, vol. 7, pp. 12 035–12 046, 2019.

[37] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous cdma systems over awgn channel," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1616–1626, 2008.

[38] Y. Du, B. Dong, P. Gao, Z. Chen, J. Fang, and S. Wang, "Low-complexity lds-cdma detection based on dynamic factor graph," in *2016 IEEE Globecom Workshops (GC Wkshps)*, vol. 67, no. 10, Dec 2016, pp. 1–6.

[39] L. Wen, "Non-orthogonal multiple access schemes for future cellular systems," Ph.D. dissertation, University of Surrey, 2016.

[40] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–5.

[41] 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 *2014 IEEE International Conference on Communication Systems*, Nov 2014, pp. 222–226.

[42] N. I. Kim and D. Cho, "Hybrid Multiple Access System Based on Non Orthogonality and Sparse Code, year=2017," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6.

[43] M. Moltafet, N. Mokari, M. R. Javan, H. Saeedi, and H. Pishro-Nik, "A New Multiple Access Technique for 5G: Power Domain Sparse Code Multiple Access (PSMA)," *IEEE Access*, vol. 6, pp. 747–759, 2018.

[44] S. Sharma, K. Deka, V. Bhatia, and A. Gupta, "Joint Power-Domain and SCMA-Based NOMA System for Downlink in 5G and Beyond," *IEEE Communications Letters*, vol. 23, no. 6, pp. 971–974, 2019.

[45] M. Mohammadkarimi, M. A. Raza, and O. A. Dobre, "Signature-based nonorthogonal massive multiple access for future wireless networks: Uplink massive connectivity for machine-type communications," *IEEE Vehicular Technology Magazine*, vol. 13, no. 4, pp. 40–50, 2018.

[46] X. Dai, Z. Zhang, B. Bai, S. Chen, and S. Sun, "Pattern division multiple access: A new multiple access technology for 5g," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 54–60, 2018.

[47] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple access—a novel nonorthogonal multiple access for fifth-generation radio networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, 2017.

[48] T. J. Myers, "Random phase multiple access communication interface system and method," 08 2010, uS Patent - Ingenu Inc. [Online]. Available: https://patents.google.com/patent/US7782926

[49] Nokia, "Performance of interleave division mul- tiple access (idma) in combination with ofdm family waveforms," document R1-165021, 3GPP TSG-RAN WG1 85, May 2016.

[50] ETRI, "Low code rate and signature based multiple access scheme for new radio," document R1-164869, TSG-RAN WG1 85, May 2016.

[51] Qualcomm, "Candidate nr multiple access schemes," document R1-163510,TSG-RAN WG1 84b, April 2016.

[52] Nokia, "Non-orthogonal multiple access for new radio," document R1-165019, 3GPP TSG-RAN WG1 85, May 2016.

[53] A. Wolfgang, S. Chen, and L. Hanzo, "Parallel interference cancellation based turbo space-time equalization in the sdma uplink," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 609–616, 2007.

[54] J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Turbo multi-user detection for ofdm/sdma systems relying on differential evolution aided iterative channel estimation," *IEEE Transactions on Communications*, vol. 60, no. 6, pp. 1621–1633, 2012.

[55] ——, "Evolutionary-algorithm-assisted joint channel estimation and turbo multiuser detection/decoding for ofdm/sdma," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1204–1222, 2014.

[56] 3GPP-36.866, "Study on network-assisted interference cancellation and suppression (NAICS) for LTE," 3rd Generation Partnership Project (3GPP), Technical report (TR), 03 2014, version 12.0.1.

[57] 3GPP-36.859, "Study on Downlink Multiuser Superposition Transmission (MUST) for LTE," 3rd Generation Partnership Project (3GPP), Technical report (TR), 04 2015, version 13.0.0.

[58] 3GPP-RP-163510, "Candidate NR multiple access schemes ," Qualcomm Incorporated , RP, 03 2017.

[59] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, 2015.

[60] 3GPP-RP-170681, " New SI: Study on 5G Non-Orthogonal Multiple Access ," ZTE Corporation, New WI/SI proposals for New Radio, 03 2017.

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

[62] B. Wang, K. Wang, Z. Lu, T. Xie, and J. Quan, "Comparison study of non-orthogonal multiple access schemes for 5g," in *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, vol. 5, no. 1, June 2015, pp. 1–5.

[63] M. Moltafet, N. M. Yamchi, M. R. Javan, and P. Azmi, "Comparison Study Between PD-NOMA and SCMA," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1830–1834, Feb. 2018.

[64] H. Nikopour and H. Baligh, "Sparse code multiple access," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013, pp. 332–336.

[65] S. Moon, H.-S. Lee, and J.-W. Lee, "SARA: Sparse Code Multiple Access-Applied Random Access for IoT Devices," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3160–3174, Aug. 2018.

[66] K. Au, L. Zhang, H. Nikopour, E. Yi, A. Bayesteh, U. Vilaipornsawai, J. Ma, and P. Zhu, "Uplink contention based SCMA for 5g radio access," in *2014 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2014, pp. 900–905.

[67] L. Liu, M. Sheng, J. Liu, X. Wang, and J. Li, "Success Probability and Area Spectral Efficiency in SCMA Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7764–7768, Aug. 2018.

[68] R. Laroia, N. Farvardin, and S. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1044–1056, Jul. 1994.

[69] M. Rebhi, K. Hassan, K. Raoof, and P. Chargé, "An adaptive uplink scma scheme based on channel state information," in *Future Networks (5G and beyond) Workshop*, Paris, France, Mar. 2020, pp. 1–7.

[70] S. Zhang, B. Xiao, K. Xiao, Z. Chen, and B. Xia, "Design and analysis of irregular sparse code multiple access," in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, 2015, pp. 1–5.

[71] L. Yu, P. Fan, Z. Ma, X. Lei, and D. Chen, "An optimized design of irregular scma codebook based on rotated angles and exit chart," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 2016, pp. 1–5.

[72] D. Zhai, "Adaptive codebook design and assignment for energy saving in scma networks," *IEEE Access*, vol. 5, pp. 23 550–23 562, 2017.

[73] S. Han, C. Guo, W. Meng, and C. Li, "A flexible resource scheduling scheme for an adaptive scma system," *Computer Networks*, vol. 129, pp. 384 – 391, 2017, special Issue on 5G Wireless Networks for IoT and Body Sensors.

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

[75] M. Beko and R. Dinis, "Designing good multi-dimensional constellations," *IEEE Wireless Communications Letters*, vol. 1, no. 3, pp. 221–224, 2012.

[76] J. Peng, W. Chen, B. Bai, X. Guo, and C. Sun, "Joint Optimization of Constellation With Mapping Matrix for SCMA Codebook Design," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 264–268, Mar. 2017.

[77] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "Scma codebook design," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, 2014, pp. 1–5.

[78] A. Bayesteh, H. Nikopour, M. Taherzadeh, H. Baligh, and J. Ma, "Low complexity techniques for scma detection," in *2015 IEEE Globecom Workshops (GC Wkshps)*, 2015, pp. 1–6.

[79] M. Taherzadehboroujeni, H. Nikopour, A. Bayesteh, and A. Baligh, "System and method for designing and using multidimensional constellations," u.S Patent 9,509,379, Nov. 2016.

[80] T. Metkarunchit, "Scma codebook design base on circular-qam," in *2017 Integrated Communications, Navigation and Surveillance Conference (ICNS)*, 2017, pp. 3E1–1–3E1–8.

[81] F. Wei and W. Chen, "Low complexity iterative receiver design for sparse code multiple access," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 621–634, 2017.

[82] Yan Xin, Zhengdao Wang, and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding," *IEEE Transactions on Wireless Communications*, vol. 2, no. 2, pp. 294–309, 2003.

[83] C. D'Amours, "Parity bit selected spreading sequences: a block coding approach to spread spectrum," *IEEE Communications Letters*, vol. 9, no. 1, pp. 16–18, 2005.

[84] M. Shi, C. D'Amours, and A. Yongacoglu, "Design of spreading permutations for mimo-cdma based on space-time block codes," *IEEE Communications Letters*, vol. 14, no. 1, pp. 36–38, 2010.

[85] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[86] S. P. Herath, N. H. Tran, and T. Le-Ngoc, "Rotated multi-d constellations in rayleigh fading: Mutual information improvement and pragmatic approach for near-capacity performance in high-rate regions," *IEEE Transactions on Communications*, vol. 60, no. 12, pp. 3694–3704, 2012.

[87] Q. Xie, Z. Yang, J. Song, and L. Hanzo, "Exit-chart-matching-aided near-capacity coded modulation design and a bicm-id design example for both gaussian and rayleigh channels," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 3, pp. 1216–1227, 2013.

[88] J. Bao, Z. Ma, Z. Ding, G. K. Karagiannidis, and Z. Zhu, "On the Design of Multiuser Codebooks for Uplink SCMA Systems," *IEEE Communications Letters*, vol. 20, no. 10, pp. 1920–1923, Oct. 2016.

[89] J. Bao, Z. Ma, M. A. Mahamadu, Z. Zhu, and D. Chen, "Spherical codes for scma codebook," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1–5.

[90] M. Vameghestahbanati, "Hypercube-based multidimensional constellation design for uplink scma systems," in *2020 IEEE International Conference on Communications (ICC)*, 2020.

[91] Altera Innovate Asia FPGA Design Contest. (2015) 5G Algorithm Innovation Competition.

[92] L. Yu, X. Lei, P. Fan, and D. Chen, "An optimized design of scma codebook based on star-qam signaling constellations," in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, 2015, pp. 1–5.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

[93] L. Yu, P. Fan, D. Cai, and Z. Ma, "Design and analysis of scma codebook based on star-qam signaling constellations," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 543–10 553, Nov 2018.

[94] J. Bao, Z. Ma, M. Xiao, T. A. Tsiftsis, and Z. Zhu, "Bit-interleaved coded scma with iterative multiuser detection: Multidimensional constellations design," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5292–5304, 2018.

[95] M. Grant and S. Boyd, " CVX: Matlab Software for Disciplined Convex Programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[96] W. Ye, W. Chen, X. Guo, C. Sun, and L. Hanzo, "Quantum search-aided multi-user detection for sparse code multiple access," *IEEE Access*, vol. 7, pp. 52 804–52 817, 2019.

[97] M. Vameghestahbanati, "Designing Multidimensional Constellations and Efficient Detection Schemes for Sparse Code Multiple Access(SCMA)Systems," Ph.D. dissertation, Carleton University, 2020.

[98] M. Alam and Q. Zhang, "Performance study of scma codebook design," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–5.

[99] C. Yan, G. Kang, and N. Zhang, "A dimension distance-based scma codebook design," *IEEE Access*, vol. 5, pp. 5471–5479, 2017.

[100] K. Lai, J. Lei, L. Wen, G. Chen, W. Li, and P. Xiao, "Secure Transmission With Randomized Constellation Rotation for Downlink Sparse Code Multiple Access System," *IEEE Access*, vol. 6, pp. 5049–5063, 2018.

[101] Z. Mheich, L. Wen, P. Xiao, and A. Maaref, "Design of scma codebooks based on golden angle modulation," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1501–1509, 2019.

[102] M. A. Imran, Y. A. Sambo, and Q. H. Abbasi, *Enabling 5G Communication Systems to Support Vertical Industries*. John Wiley & Sons Ltd, 2019.

[103] S. Zhang, K. Xiao, B. Xiao, Z. Chen, B. Xia, D. Chen, and S. Ma, "A capacity-based codebook design method for sparse code multiple access systems," in *2016 8th International Conference on Wireless Communications Signal Processing (WCSP)*, 2016, pp. 1–5.

[104] Y. Zhou, Q. Yu, W. Meng, and C. Li, "Scma codebook design based on constellation rotation," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[105] K. Xiao, B. Xia, Z. Chen, B. Xiao, D. Chen, and S. Ma, "On Capacity-Based Codebook Design and Advanced Decoding for Sparse Code Multiple Access Systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3834–3849, Jun. 2018.

[106] M. Kulhandjian and C. D'Amours, "Design of permutation-based sparse code multiple access system," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–6.

[107] F. R. Kschischang, B. J. Frey, and H. . Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[108] Y. Qi, G. Wu, S. Hu, and Y. Gao, "Parallel-implemented message passing algorithm for scma decoder based on gpgpu," in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2017, pp. 1–6.

[109] W. B. Ameur, P. Mary, M. Dumay, J. Hélard, and J. Schwoerer, "Performance study of mpa, log-mpa and max-log-mpa for an uplink scma scenario," in *2019 26th International Conference on Telecommunications (ICT)*, 2019, pp. 411–416.

[110] K. Xiao, B. Xiao, S. Zhang, Z. Chen, and B. Xia, "Simplified multiuser detection for scma with sum-product algorithm," in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, 2015, pp. 1–5.

[111] J. Dai, K. Niu, C. Dong, and J. Lin, "Improved Message Passing Algorithms for Sparse Code Multiple Access," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 9986–9999, Nov. 2017.

[112] Jia Zou, Hui Zhao, and Wenxiu Zhao, "Low-complexity interference cancellation receiver for sparse code multiple access," in *2015 IEEE 6th International Symposium on Microwave, Antenna, Propagation, and EMC Technologies (MAPE)*, 2015, pp. 277–282.

[113] P. Gao, Y. Du, B. Dong, W. Zhu, Z. Chen, and X. Wang, "Low-Complexity CS-Aided MPA Detector for SCMA Systems," *IEEE Communications Letters*, vol. 22, no. 4, pp. 784–787, Apr. 2018.

[114] S. Han, Y. Huang, and B. Wang, "A method for analysing and improving the multi-user detection algorithm of scma," in *Wireless Internet*. Springer International Publishing, 2018, pp. 103–115.

[115] L. Ma, S. Tong, H. Zheng, B. Bai, and X. Dai, "Edgewise serial message passing detection of uplink scma systems for better user fairness and faster convergence rate," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1285–1288, 2019.

[116] L. Yang, Y. Liu, and Y. Siu, "Low complexity message passing algorithm for scma system," *IEEE Communications Letters*, vol. 20, no. 12, pp. 2466–2469, 2016.

[117] G. Zhang, Z. Gu, Q. Zhao, J. Ren, and W. Lu, "A threshold-based max-log-mpa low complexity multiuser detection algorithm," *Sensors*, vol. 20, no. 4, p. 1016, 2020.

[118] H. Wu, X. Xiong, and H. Gao, "Low complexity detection based on selective message passing for scma systems," *Electronics Letters*, vol. 54, no. 8, pp. 533–535, 2018.

[119] B. F. d. Silva, D. Le Ruyet, and B. F. Uchoa-Filho, "Threshold-based edge selection mpa for scma," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 2957–2966, 2020.

[120] C. Zhang, C. Yang, X. Pang, W. Song, W. Xu, S. Zhang, Z. Zhang, and X. You, "Efficient sparse code multiple access decoder based on deterministic message passing algorithm," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 3562–3574, 2020.

[121] M. Jia, L. Wang, Q. Guo, X. Gu, and W. Xiang, "A low complexity detection algorithm for fixed up-link scma system in mission critical scenario," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3289–3297, 2018.

[122] K. Lai, L. Wen, J. Lei, J. Zhong, G. Chen, and X. Zhou, "Simplified sparse code multiple access receiver by using truncated messages," *IET Communications*, vol. 12, no. 16, pp. 1937–1945, 2018.

[123] X. Peng, Z. Pan, K. Lai, L. Wen, and J. Lei, "Low complexity receiver of sparse code multiple access based on dynamic trellis," *IET Communications*, vol. 14, no. 9, pp. 1420–1427, 2020.

[124] X. Shi, P. Ren, and D. Xu, "A flexible iterative log-mpa detector for uplink scma systems," in *IoT as a Service*. Springer International Publishing, 2019, pp. 292–302.

[125] F. Bellili, F. Sohrabi, and W. Yu, "Generalized approximate message passing for massive mimo mmwave channel estimation with laplacian prior," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3205–3219, 2019.

[126] Y. Huang, Y. Li, M. Zhao, X. Xu, and J. Wang, "Two Simplified Multiuser Detection Algorithms for Uplink SCMA Systems via Generalized Approximate Message Passing," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–5.

[127] G. Kang, H. M. Kirn, Y. Shin, and O. Shin, "Multiple antenna mpa detection for uplink scma systems based on selective gaussian approximation," in *2017 23rd Asia-Pacific Conference on Communications (APCC)*, 2017, pp. 1–4.

[128] S.-C. Lim and H. Park, "Gaussian Approximation-Based Belief Propagation Receiver for Uplink SCMA Systems With Imperfect CSIR," *IEEE Communications Letters*, vol. 22, no. 12, pp. 2611–2614, Dec. 2018.

[129] J. Dai, K. Niu, and J. Lin, "Iterative Gaussian-Approximated Message Passing Receiver for MIMO-SCMA System," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 753–765, Jun. 2019.

[130] X. Meng, Y. Wu, Y. Chen, and M. Cheng, "Low Complexity Receiver for Uplink SCMA System via Expectation Propagation," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. San Francisco, CA, USA: IEEE, Mar. 2017, pp. 1–5.

[131] W. Yuan, N. Wu, Q. Guo, Y. Li, C. Xing, and J. Kuang, "Iterative receivers for downlink mimo-scma: Message passing and distributed cooperative detection," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3444–3458, 2018.

[132] J. Chen, Z. Zhang, S. Fu, and J. Hu, "A Joint Update Parallel MCMC-Method-Based Sparse Code Multiple Access Decoder," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1280–1291, Feb. 2018.

[133] P. Wang, L. Liu, S. Zhou, G. Peng, S. Yin, and S. Wei, "Near-optimal mimo-scma uplink detection with low-complexity expectation propagation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 1025–1037, 2020.

[134] T. Minka, "Low Complexity Receiver for Uplink SCMA System via Expectation Propagation," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence (UAI)*, August 2001, pp. 362–369.

[135] Q. Yu, H. Li, W. Meng, and W. Xiang, "Sparse code multiple access asynchronous uplink multiuser detection algorithm," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5557–5569, 2019.

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/OJCOMS.2021.3081166, IEEE Open Journal of the Communications Society

[137] S. Chen, W. Wang, D. Zhang, and X. Gao, "Robust Approximate Message Passing Detection Based on Minimizing Bethe Free Energy for Massive MIMO Systems," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–6.

[138] D. Zhang, W. Wang, G. P. Fettweis, and X. Gao, "Unifying message passing algorithms under the framework of constrained bethe free energy minimization," *CoRR*, vol. 1703.10932, 2017.

[139] W. Yuan, N. Wu, C. Yan, Y. Li, X. Huang, and L. Hanzo, "A low-complexity energy-minimization-based scma detector and its convergence analysis," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 12 398–12 403, 2018.

[140] C. Zhang, Y. Luo, and Y. Chen, "A Low Complexity Detection Algorithm for SCMA," Nov. 2016.

[141] C. Zhang, Y. Luo, and Y. Chen, "A low-complexity scma detector based on discretization," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2333–2345, 2018.

[142] C. Hung and T. Sang, "A sphere decoding algorithm for mimo channels," in *2006 IEEE International Symposium on Signal Processing and Information Technology*, 2006, pp. 502–506.

[143] F. Wei and W. Chen, "A low complexity scma decoder based on list sphere decoding," in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[144] L. Tian, M. Zhao, J. Zhong, P. Xiao, and L. Wen, "A low complexity detector for downlink scma systems," *IET Communications*, vol. 11, no. 16, pp. 2433–2439, 2017.

[145] L. Yang, X. Ma, and Y. Siu, "Low Complexity MPA Detector Based on Sphere Decoding for SCMA," *IEEE Communications Letters*, vol. 21, no. 8, pp. 1855–1858, Aug. 2017.

[146] M. Vameghestahbanati, E. Bedeer, I. Marsland, R. H. Gohary, and H. Yanikomeroglu, "Enabling Sphere Decoding for SCMA," *IEEE Communications Letters*, vol. 21, no. 12, pp. 2750–2753, Dec. 2017.

[147] G. Chen, J. Dai, K. Niu, and C. Dong, "Optimal receiver design for SCMA system," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct. 2017, pp. 1–6, iSSN: 2166-9589.

[148] G. Chen, J. Dai, K. Niu, and C. Dong, "Sparsity-inspired sphere decoding (si-sd): A novel blind detection algorithm for uplink grant-free sparse code multiple access," *IEEE Access*, vol. 5, pp. 19 983–19 993, 2017.

[149] Y. Han, W. Zhou, M. Zhao, and S. Zhou, "Enabling High Order SCMA Systems in Downlink Scenarios With a Serial Coding Scheme," *IEEE Access*, vol. 6, pp. 33 796–33 809, 2018.

[150] L. Li, J. Wen, X. Tang, and C. Tellambura, "Modified Sphere Decoding for Sparse Code Multiple Access," *IEEE Communications Letters*, vol. 22, no. 8, pp. 1544–1547, Aug. 2018.

[151] M. Vameghestahbanati, I. Marsland, R. H. Gohary, and H. Yanikomeroglu, "A novel sd-based detection for generalized scma constellations," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10 278–10 282, 2019.

[152] P. Botsinis, D. Alanis, Z. Babar, H. V. Nguyen, D. Chandra, S. X. Ng, and L. Hanzo, "Quantum search algorithms for wireless communications," *IEEE Communications Surveys Tutorials*, vol. 21, no. 2, pp. 1209–1242, 2019.

[153] D. Hui, S. Sandberg, Y. Blankenship, M. Andersson, and L. Grosjean, "Channel coding in 5g new radio: A tutorial overview and performance comparison with 4g lte," *IEEE Vehicular Technology Magazine*, vol. 13, no. 4, pp. 60–69, 2018.

[154] J. H. Bae, A. Abotabl, H.-P. Lin, K.-B. Song, and J. Lee, "An overview of channel coding for 5g nr cellular communications," *APSIPA Transactions on Signal and Information Processing*, vol. 8, p. e17, 2019.

[155] K. Lai, L. Wen, J. Lei, P. Xiao, A. Maaref, and M. A. Imran, "Sub-Graph Based Joint Sparse Graph for Sparse Code Multiple Access Systems," *IEEE Access*, vol. 6, pp. 25 066–25 080, 2018.

[156] W. Sun, Y. Su, Y. Ueng, and C. Yang, "An ldpc-coded scma receiver with multi-user iterative detection and decoding," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 9, pp. 3571–3584, 2019.

[157] H. Mu, Y. Tang, L. Li, Z. Ma, P. Fan, and W. Xu, "Polar coded iterative multiuser detection for sparse code multiple access system," *China Communications*, vol. 15, no. 11, pp. 51–61, 2018.

[158] Z. Pan, E. Li, L. Zhang, J. Lei, and C. Tang, "Design and optimization of joint iterative detection and decoding receiver for uplink polar coded scma system," *IEEE Access*, vol. 6, pp. 52 014–52 026, 2018.

[159] J. Jiao, K. Liang, B. Feng, Y. Wang, S. Wu, and Q. Zhang, "Joint channel estimation and decoding for polar coded scma system over fading channels," *IEEE Transactions on Cognitive Communications and Networking*, pp. 1–1, 2020.

[160] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[161] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep Learning for an Effective Non-orthogonal Multiple Access Scheme," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8440–8450, 2018.

[162] C. Lin, Q. Chang, and X. Li, " A Deep Learning Approach for MIMO-NOMA Downlink Signal Detection," *Sensors*, vol. 19, no. 11, 2019.

[163] J. Luo, J. Tang, D. K. C. So, G. Chen, K. Cumanan, and J. A. Chambers, "A Deep Learning-Based Approach to Power Minimization in Multi-Carrier NOMA With SWIPT," *IEEE Access*, vol. 7, pp. 17 450–17 460, 2019.

[164] M. Liu, T. Song, and G. Gui, "Deep Cognitive Perspective: Resource Allocation for NOMA-Based Heterogeneous IoT With Imperfect SIC," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2885–2894, 2019.

[165] N. Ye, X. Li, H. Yu, A. Wang, W. Liu, and X. Hou, "Deep Learning Aided Grant-Free NOMA Toward Reliable Low-Latency Access in Tactile Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2995–3005, 2019.

[166] M. K. Hasan, M. Shahjalal, M. M. Islam, M. M. Alam, M. F. Ahmed, and Y. M. Jang, "The Role of Deep Learning in NOMA for 5G and Beyond Communications," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 303–307.

[167] J. M. Kang, I. M. Kim, and C. J. Chun, "Deep Learning-Based MIMO-NOMA With Imperfect SIC Decoding," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3414–3417, 2020.

[168] K. Chitti, J. Vieira, and M. Behrooz, "Deep-Learning based Multiuser Detection for NOMA," 11 2020.

[169] M. Kim, N. Kim, W. Lee, and D. Cho, "Deep learning-aided scma," *IEEE Communications Letters*, vol. 22, no. 4, pp. 720–723, 2018.

[170] F. Sun, K. Niu, and C. Dong, "Deep learning based joint detection and decoding of non-orthogonal multiple access systems," in *2018 IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–5.

[171] C. Lu, W. Xu, H. Shen, H. Zhang, and X. You, "An enhanced SCMA detector enabled by deep neural network," *CoRR*, vol. abs/1808.08015, 2018.

[172] Y. Han, Z. Wang, Q. Guo, and W. Xiang, "Deep learning-based detection for moderate-density code multiple access in iot networks," *IEEE Communications Letters*, vol. 24, no. 1, pp. 122–125, 2020.

[173] S. Sharma and Y. Hong, "A hybrid multiple access scheme via deep learning-based detection," *IEEE Systems Journal*, pp. 1–4, 2020.

[174] J. Lin, S. Feng, Y. Zhang, Z. Yang, and Y. Zhang, "A novel deep neural network based approach for sparse code multiple access," *Neurocomputing*, vol. 382, pp. 52–63, 2020.

[175] G. Kang, H. M. Kim, Y. Shin, and O. Shin, "Message passing algorithm based on qr decomposition for an scma system with multiple antennas," in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, 2017, pp. 941–944.

[176] Z. Wu, C. Zhang, X. Shen, and H. Jiao, "Low complexity uplink sfbc-based mimo-scma joint decoding algorithm," in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, 2017, pp. 968–972.

[177] J. Dai, G. Chen, K. Niu, and J. Lin, "Partially Active Message Passing Receiver for MIMO-SCMA Systems," *IEEE Wireless Communications Letters*, vol. 7, no. 2, pp. 222–225, Apr. 2018.

[178] L. Tian, M. Zhao, J. Zhong, and L. Wen, "Resource-selection based low complexity detector for uplink scma systems with multiple antennas," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 316–319, 2018.

[179] A. Kiani and N. Ansari, "Edge computing aware noma for 5g networks," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1299–1306, 2018.

[180] Q. Fan and N. Ansari, "Application aware workload allocation for edge computing-based iot," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2146–2153, 2018.

[181] Y. Gu, Z. Chang, M. Pan, L. Song, and Z. Han, "Joint radio and computational resource allocation in iot fog computing," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7475–7484, 2018.

[182] J. Wang, C. Zhang, R. Li, G. Wang, and J. Wang, "Narrow-band scma: A new solution for 5g iot uplink communications," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 2016, pp. 1–5.

[183] L. Miuccio, D. Panno, and S. Riolo, "Joint congestion control and resource allocation for massive mtc in 5g networks based on scma," in *2019 15th International Conference on Telecommunications (ConTEL)*, 2019, pp. 1–8.

[184] I. Hidayat, L. Meylani, A. Kurniawan, M. S. Arifianto, and K. Anwar, "Doubly irregular sparse code multiple access with exit analysis," in *2018 International Conference on Signals and Systems (ICSigSys)*, 2018, pp. 115–119.

[185] A. Alnoman, S. Erkucuk, and A. Anpalagan, "Sparse code multiple access-based edge computing for iot systems," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 7152–7161, 2019.

[186] J. Cui, P. Fan, X. Lei, Z. Ma, and Z. Ding, "Downlink Power Allocation in SCMA with Finite-Alphabet Constraints," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, Jun. 2017, pp. 1–5.

[187] M. Dabiri and H. Saeedi, "Dynamic scma codebook assignment methods: A comparative study," *IEEE Communications Letters*, vol. 22, no. 2, pp. 364–367, 2018.

[188] W. Liqiang, L. Xinyue, G. Shan, and H. Xingling, "Resource allocation strategy in SCMA based multiple access system," in *2017 First International Conference on Electronics Instrumentation Information Systems (EIIS)*, Jun. 2017, pp. 1–6.

[189] J. Zhang, M. Zhang, D. Li, Y. Yang, and Y. Zhao, "Resource allocation for downlink scma system based on coalitional game," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018, pp. 726–731.

[190] S. Han, Y. Huang, W. Meng, C. Li, N. Xu, and D. Chen, "Optimal power allocation for scma downlink systems based on maximum capacity," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1480–1489, 2019.

[191] Wei Yu, Wonjong Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, 2004.

[192] Z. Li, W. Chen, F. Wei, F. Wang, X. Xu, and Y. Chen, "Joint codebook assignment and power allocation for scma based on capacity with gaussian input," in *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, 2016, pp. 1–6.

[193] Y. Li, M. Sheng, Z. Sun, Y. Sun, L. Liu, D. Zhai, and J. Li, "Cost-efficient codebook assignment and power allocation for energy efficiency maximization in scma networks," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 2016, pp. 1–5.

[194] Q. Yuan, Y. Dai, B. Zhu, and J. Song, "A resource allocation method based on energy efficiency in scma systems," in *2019 IEEE 2nd International Conference on Electronics Technology (ICET)*, 2019, pp. 34–39.

[195] B. Fontana da Silva, D. Silva, B. F. Uchôa-Filho, and D. Le Ruyet, "A Multistage Method for SCMA Codebook Design Based on MDS Codes," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1524–1527, 2019.

[196] P. P. Waghmare, "An optimized code book design and assignment based on 32-QAM constellation in downlink SCMA systems," *Digital Signal Processing*, vol. 109, p. 102919, 2021.

[197] Fang Shu, Gang Wu, and Shao-qian Li, "Dual codebook based multi-user MIMO precoding and scheduling scheme with limited feedback," in *2009 International Conference on Communications, Circuits and Systems*, 2009, pp. 19–23.

[198] J. Liu, G. Wu, S. Li, and O. Tirkkonen, "On fixed-point implementation of log-mpa for scma signals," *IEEE Wireless Communications Letters*, vol. 5, no. 3, pp. 324–327, 2016.

[199] C. Yang, C. Zhang, S. Zhang, and X. You, "Efficient hardware architecture of deterministic mpa decoder for scma," in *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2016, pp. 293–296.

[200] K. Han, J. Hu, J. Chen, and H. Lu, "A low complexity sparse code multiple access detector based on stochastic computing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 2, pp. 769–782, 2018.

[201] T. Rappaport, R. Heath, R. Daniels, and J. Murdock, *Millimeter wave wireless communications*. Prentice Hall, 2015.

[202] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave mimo systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[203] H. Hosseini, A. Anpalagan, K. Raahemifar, and S. Erkucuk, "Wavelet-based cognitive scma system for mmwave 5g communication networks," *IET Communications*, vol. 11, no. 6, pp. 831–836, 2017.

[204] S. Tang, Z. Ma, M. Xiao, and L. Hao, "Performance analysis for mmwave mimo-scma systems using lens antenna array," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–5.

[205] ——, "Hybrid transceiver design for beamspace mimo-noma in code-domain for mmwave communication using lens antenna array," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2118–2127, 2020.

[206] H. Baligh, A. Bayesteh, Y. Lin, U. Vilaipornsawai, and K. Zarifi, "Two-tier distributed and open loop multi-point cooperation using scma," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016, pp. 1–5.

[207] Lin Luo, Lihua Li, and Xin Su, "Optimization of resource allocation in relay assisted multi-user scma uplink network," in *2017 International Conference on Computing, Networking and Communications (ICNC)*, 2017, pp. 282–286.

[208] Y. Zhang, X. Wang, D. Wang, Y. Lan, and Y. Zhang, "Outage performance of relay sparse code multiple access networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1–6.

[209] R. Raut, R. Sawant, and S. Madbushi, *Sparse Code Multiple Access (SCMA) for Cognitive Radio*, April 2020, pp. 185–194.

[210] G. K. Shekhawat and R. P. Yadav, "Sparse code multiple access based cooperative spectrum sensing in 5g cognitive radio networks," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, 2020, pp. 1–6.

[211] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1550–1573, 2014.

[212] Z. Li, S. Han, and X. Wang, "Power allocation scheme for physical-layer security of two-way untrusted relay in scma networks," in *2020 International Conference on Computing, Networking and Communications (ICNC)*, 2020, pp. 497–501.

[213] K. Lai, J. Lei, L. Wen, G. Chen, W. Li, and P. Xiao, "Secure transmission with randomized constellation rotation for downlink sparse code multiple access system," *IEEE Access*, vol. 6, pp. 5049–5063, 2018.

[214] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy harvesting wireless communications: A review of recent advances," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 360–381, 2015.

[215] M. Abedi, N. Mokari, M. Reza Javan, and E. A. Jorswieck, "Single or multiple frames content delivery for next-generation networks?" *IEEE Access*, vol. 7, pp. 152 501–152 521, 2019.

[216] M. Wen, B. Zheng, K. J. Kim, M. Di Renzo, T. A. Tsiftsis, K. Chen, and N. Al-Dhahir, "A Survey on Spatial Modulation in Emerging Wireless Systems: Research Progresses and Applications," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 9, pp. 1949–1972, 2019.

[217] Y. Liu, L. Yang, and L. Hanzo, "Spatial Modulation Aided Sparse Code-Division Multiple Access," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1474–1487, 2018.

[218] Z. Pan, J. Luo, J. Lei, L. Wen, and C. Tang, "Uplink Spatial Modulation SCMA System," *IEEE Communications Letters*, vol. 23, no. 1, pp. 184–187, 2019.

[219] I. Al-Nahhal, O. A. Dobre, E. Basar, and S. Ikki, "Low-Cost Uplink Sparse Code Multiple Access for Spatial Modulation," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 9313–9317, 2019.

[220] I. Al-Nahhal, O. A. Dobre, and S. Ikki, "On the Complexity Reduction of Uplink Sparse Code Multiple Access for Spatial Modulation," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 6962–6974, 2020.

[221] P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible light communication, networking, and sensing: A survey, potential and challenges," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2047–2077, 2015.

[222] J. An and W. Chung, "Single-led multichannel optical transmission with scma for long range health information monitoring," *Journal of Lightwave Technology*, vol. 36, no. 23, pp. 5470–5480, 2018.

[223] S. Lou, C. Gong, Q. Gao, and Z. Xu, "Scma with low complexity symmetric codebook design for visible light communication," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6.

[224] B. Lin, C. Wei, J. Xu, Q. Lai, X. Tang, and X. Shen, "Experimental demonstration of pd-scma for visible light communications," in *2019 Asia Communications and Photonics Conference (ACP)*, 2019, pp. 1–3.

[225] Q. Gao, S. Hu, C. Gong, E. Serpedin, K. Qaraqe, and Z. Xu, "Distance-range-oriented constellation design for vlc-scma downlink with signal-dependent noise," *IEEE Communications Letters*, vol. 23, no. 3, pp. 434–437, 2019.

[226] R. Mitra, S. Sharma, G. Kaddoum, and V. Bhatia, "Color-domain scma noma for visible light communication," *IEEE Communications Letters*, pp. 1–1, 2020.

[227] B. Lin, X. Tang, and Z. Ghassemlooy, "A power domain sparse code multiple access scheme for visible light communications," *IEEE Wireless Communications Letters*, vol. 9, no. 1, pp. 61–64, 2020.

[228] L. Yang, X. Lin, X. Ma, and K. Song, "Clipping noise-aided message passing algorithm for scma-ofdm system," *IEEE Communications Letters*, vol. 22, no. 10, pp. 2156–2159, 2018.

[229] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless Communications Through Reconfigurable Intelligent Surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[230] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, and L. Hanzo, "Reconfigurable Intelligent Surface Aided NOMA Networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2575–2588, 2020.

[231] G. Yang, X. Xu, Y. C. Liang, and M. Di Renzo, "Reconfigurable Intelligent Surface Assisted Non-Orthogonal Multiple Access," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2021.

[232] I. Al-Nahhal, O. A. Dobre, and E. Basar, "Reconfigurable Intelligent Surface-Assisted Uplink Sparse Code Multiple Access," *IEEE Communications Letters*, pp. 1–1, 2021.