Unsourced Multiple Access for 6G Massive Machine Type Communications

Yuanjie Li[1], Jincheng Dai[1], Zhongwei Si[1,*], Kai Niu[1], Chao Dong[1], Jiaru Lin[1], Sen Wang[2], Yifei Yuan[2]

[1] Key Laboratory of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2] China Mobile Research Institute, Beijing 100053, China
[*] The corresponding author, email: sizhongwei@bupt.edu.cn

Abstract: Unsourced multiple access (UMA) is a multi-access technology for massive, low-power, uncoordinated, and unsourced Machine Type Communication (MTC) networks. It ensures transmission reliability under the premise of high energy efficiency. Based on the analysis of the 6G MTC key performance indicators (KPIs) and scenario characteristics, this paper summarizes its requirements for radio access networks. Following this, the existing multiple access models are analyzed under these standards to determine UMA's advantages for 6G MTC according to its design characteristics. The critical technology of UMA is the design of its multiple-access coding scheme. Therefore, the existing UMA coding schemes from different coding paradigms are further summarized and compared. In particular, this paper comprehensively considers the energy efficiency and computational complexity of these schemes, studies the changes of the above two indexes with the increase of access scale, and considers the trade-off between the two. It is revealed by the above analysis that some guiding rules of UMA coding design. Finally, the open problems and potentials in this field are given for future research.

Keywords: unsourced multiple access; machine type communications; 6G; massive random access; uncoordinated

Received: Aug. 06, 2021
Revised: Dec. 27, 2021
Editor: Yongming Huang

I. INTRODUCTION

With the freeze of 3GPP Rel-16 [1], 5G technology has entered a relatively mature stage. At this time, both academia and industry are looking forward to the 6G visions and potential technologies [2, 3]. Among them, MTC (Machine Type Communication) is a critical scenario in the consensus of the evolution from 5G to 6G. Compared with 5G, 6G MTC network puts forward very challenging KPIs (Key Performance Indicators) [4]. Under the short packet transmission setup, the requirements of reduced delay, increased connection density, and improved connection reliability will bring challenges to the existing multiple access technology, associated with the transmission model, scheduling mechanism, and resource management. Therefore, multiple access technologies need breakthroughs to meet the higher standard of 6G future networks.

Unsourced multiple access (UMA) is proposed in this context, which satisfies the evolution trend of the mobile communication system. It improves efficiency by removing coordination center. At the same time, superposition coding is used to design the unsourced codebook to counter the increased multi-user interference. In section 2.1, the analysis, KPIs, and prospects of 6G MTC application scenarios are summarized in detail. Meanwhile, the existing multiple access modes are classified into three categories in section 2.2: coordinated, grant-free, and unsourced multiple access. Based on its design principle and characteristics, the competitive advantages of UMA in the 6G MTC network compared with the other two modes are ex-

plained.

In terms of surveys in this area, [5] discussed the technical background and basic concepts of UMA. In the massive access scenario, coordinated access is a comparison scheme, and the theoretical foundation and basic design scheme are introduced. However, since Polyanskiy proposed the basic concept [6] and basic coding scheme [7] in 2017, there have been rich related research results in this field. Especially in the terms of coding schemes design, some advanced and innovative schemes have been proposed recently. Therefore, it's necessary to make a comprehensive summary of the existing UMA coding schemes and figure out some basic paradigms, ideas, and criteria of design. Section 3.2 is based on such considerations in the comparative analysis. At the same time, we sorted out the existing coding paradigms in the table and introduced in detail the encoding and decoding structures of the three state-of-art schemes in section 3.2 for further analysis and comparison.

The basic evaluation criterion for UMA coding schemes is that the minimum required SNR (Signal-to-Noise Ratio),i.e. SNR threshold, under the increase of the number of active users when the packet loss rate and the total access frame length are guaranteed. Most of the works are also aimed at designing coding schemes as close to the performance world as possible. Therefore, in section 3.3, this paper makes a comprehensive comparison of curves of the energy efficiency with the number of users of existing solutions, analyzes the advantages and disadvantages of various coding patterns, and how to approach the performance world through coding optimization. However, the complexity problem has been less analyzed, but it has an important impact on the implementation and delay of processing. Therefore, this paper constructs the complexity analysis framework of UMA coding mode takes FLOPf (float point operations per frame) as the basic merit, compares the complexity difference of different schemes under typical setups, and obtains the trade-off between them and the system energy efficiency index.

Based on the above analysis, we summarize underachieved in terms of the performance and complexity objectives of the existing solutions as a guide for future research. At the same time, the coding scheme design in fading channel and the asynchronous condition is not enough, which is very important for the practical application of UMA. The combination of MIMO and UMA will also be a rich and potential research site. Finally, as UMA simplifies the design of empty ports, its security needs to be solved by some physical layer security technologies.

II. SCENARIO AND TECHNOLOGY BACKGROUND

2.1 Scenario Evolution of Machine Type Communication

Since 5G, MTC has joined in the evolution process of the mobile communication system as a parallel application scenario with URLLC and eMBB (enhanced Mobile BroadBand) [8]. MTC mainly focuses on human-machine or machine-machine communication [9] services in industrial or social scenarios, generally transmitting interactive control information and data report of sensors [10]. The standardization of NB-IoT (NarrowBand-Internet of Things) technology [11] has been realized in 5G. With the gradual maturation of 5G MTC, 6G MTC sets higher KPIs and has richer vertical application scenarios, which can be regarded as the individual evolution or joint fusion among the three typical 5G scenarios [12].

According to [13], the 6G MTC network mainly has three types access flows: Massive MTC traffic [4], vehicle to anything (V2X) traffic [14, 15], and Industry 4.0 [16] interactive information. Among them, the concept of massive MTC Traffic is the transition from m-MTC (massive Machine Type Communications) in the 5G era to the um-MTC (ultra massive-Machine Type Communications) scenario [4], with increasing number of devices in the MTC network. Connection density increases from 10^6 to 10^7 connections/km^2 [17]. In the future, there will be vast sensors, wearable and automotive devices adding to MTC networks [18], with generally the payload of 40 to 120 bytes per packet [19]. It's essential to adapt to the carrying demand of pervasive connectivity. V2X network corresponds to the evolution from URLLC (Ultra-Reliable Low Latency Communications) to SURLLC(Super Ultra-Reabliable Low Latency Comminications) [20]. Intelligent medical applications and information transmission of the Internet of vehicles put forward higher demands in terms of high reliability and immediacy. Air-interface latency needs to be reduced from

$1ms$ to $0.1ms$ [21], and the reliability and accessibility of devices in the network needs to be improved from 99.999% to 99.9999% [22] to guarantee the high reliability and precision control interaction of rapid response between guarantor and machine or between machine and machine [23]. On the premise of deep integration of intelligence, the industrial 4.0 network should realize multiple requirements of real-time transmission, control, calculation, and interaction, as well as automatic robot production line and super-large scale sensor control network [24]. Therefore, the URLLC network needs to evolve to MURLLC (massive URLLC) [19], that is, on the premise of ensuring low latency and high reliability of data transmission, connecting to a large number of industrial equipments.In addition to these traditional KPI evolutions, the 6G MTC network has many new KPIs [13]. Considering the battery life and capacity of the devices, the energy consumption needs to be controlled at ultra-low level. And the connction set-up time is under $1ms$ [25]. For Instance, The scale of the network will be more flexible, and the access process will be more flexible and diverse. Furthermore, 6G MTC will pay more attention to the segmentation needs of vertical fields [26].

By comprehensively considering the above network characteristics and KPIs, we can summarize the specific requirements of the 6G MTC network for multiple access technology. First, the MA technology should adapt to the large-scale uplink burst short packet simultaneous transmission demand. The access load is high, and the base station needs to deal with strengthened multi-user interference in limited resources. Secondly, the air delay of the network needs to be under control, and the transmission process to be kept concise. In delay-sensitive scenarios, excessive signaling interactions may reduce transmission efficiency. The network should support unauthorized or passive access to terminals. At the same time, the packet loss rate of the system should be effectively controlled. Sensor and industrial control networks may not have higher data rate requirements for communication links than high-bandwidth high-speed communication scenarios but demand stable and low-error transmission. Finally, the MA technology has to maintain low power consumption with low computational complexity at the transmitter end. Many devices in the MTC network have limited battery capacity and computing power, which determines that the processing at the transmitter side is not too complex and frequent. In addition, on-demand access should be supported so that the devices can turn to idle state with ultra-low power consumption when not transmitting data, and reliable transmission can be achieved with low power to ensure high energy efficiency.

2.2 Multiple Access Techniques for MTC

In this subsection, we will compare and analyze three existing multiple access modes, cooperative multiple access, grant-free access, and unsourced multiple access, and show the evolution trend of access technology in the development of mobile communication systems. In particular, we will focus on the UMA in agreement with the requirements of the 6G MTC communication scenario summarized in the previous section, and explain its advantages and prospects compared with the other two. The comparison of the three multiple access models is depicted in figure 1.

2.2.1 Coordinated Multiple Access

Coordinated multiple access (CMA) is the mainstream multiple access mode in 5G and before. From 1G to 4G, FDMA (Frequency Division Multiple Access), TDMA (Time Division Multiple Access), Time Division Multiple Access, CDMA (Code Division Multiple Access), and OFDMA (Orthogonal Frequency Division Multiple Access, Orthogonal frequency division multiple access (OFDM) these are the core multiple access technologies of the ages [27, 28]. They all belong to the large category of Orthogonal Multiple Access (OMA). When it comes to 5G, non-orthogonal multiple access (NOMA) enhances the capacity region boundary [29]. Different NOMA technologies such as power domain NOMA [30], code domain NOMA [31], bit domain NOMA [32] are developed with different approaches of multiplexing.

The core idea of CMA is to utilize resource allocation, power control, or codebook adjustment among users through the cooperative process to maximize the sum rate of users according to the quality of subchannels. This function is accomplished by the coordination center at the base station side. Thus, every time the devices start the data transmission with standard RACH (Random Access CHannel) procedure exchanging signaling with coordination center, including

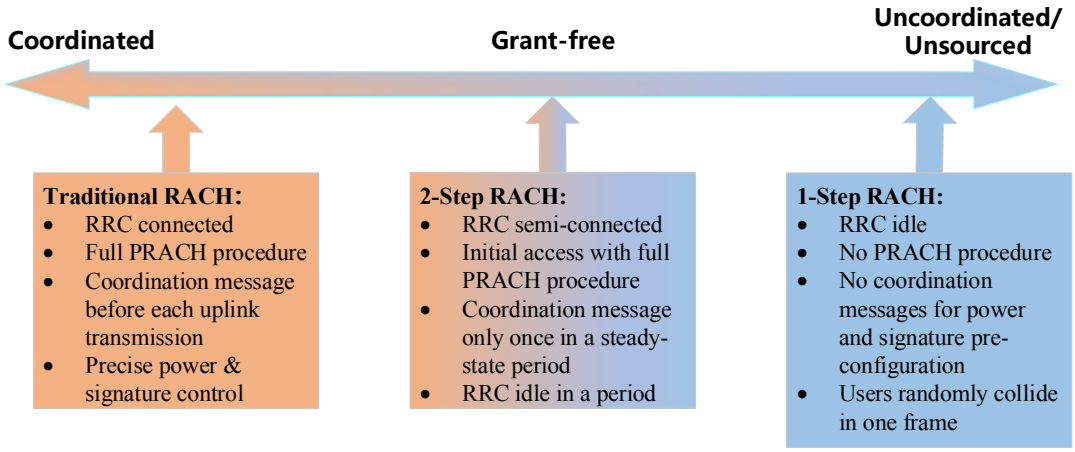


Figure 1. *RACH features of coordinated, grant-free and unsourced multiple access.*

preamble based grant acquisition and radio resource control connection (RRC) setup [33].

2.2.2 Grant-Free Multiple Access

Concerning the delay and transmission efficiency, if the RACH process-based CMA is used in the MTC scenario, transmission delay caused by exchanging signaling, processing delay of decoding and coordination signals makes it difficult for the system to meet the strict air interface latency requirement. Grant-free access is an MA method to modify the traditional RACH grant process and reduce the overhead cost [34]. Specifically, it can be implemented in two ways. One of them only performs the complete RACH process at the initial access stage but does not interact with the RRC messages during the consecutive transmissions [35]. That is, the dynamic resource allocation process is canceled, adapting semi-static RRC. Therefore, it supports both OMA and NOMA modes, and NB-IoT can also turn to the grant-free style.

The other one is a step further with a RACH-less style. The coordination center only informs the user of the available resource configuration set, and the user directly selects one of them for uplink transmissionḂecause the choice of resources from different users may collide, the NOMA approach is the only option under this model. [34] introduces MA signature [36] and Compressed Sensing (CS) [37] based methods directly evolving from corresponding coordinated NOMA technologies. The coordination center allocates optimized signature sets and spread spectrum sequence sets, and the BS separates collided users by blind detection and decoding.

2.2.3 Unsourced Multiple Access

When the MTC scenario and KPI requirements evolve to 6G (as described in section 2.1), the access models based on traditional NOMA may encounter the bottleneck. In theory, Polyanskiy pointed out in [6] that under the sharp increase of active users, the average channel capacity of users converges to zero even though the sum rate is increasing. In other words, the CMA and NOMA based on sum-rate optimization can not guarantee the connection reliability of each user. On the other hand, although the traditional NOMA technologies support the 2-step RACH in grant-free access, it is challenging to meet the low delay short packet transmission requirements due to the signaling overhead and delay caused by the coordination process. So the grant-free modification of CMA is still not enough for 6G MTC networks.

As shown in figure 2, the UMA changes grant-free to uncoordinated access. The coordination center, together with the resource allocation process, is removed. All users use a common codebook to send their packets randomly, and the collision occurs randomly in the channel. It only matters whether the packets are decoded correctly while the receiver ignores the source of them. Therefore, the optimization goal of the system changes from sum-rate to Per-User

Probability of Error (PUPE), i.e. the average packet loss rate of each user. As a result, the user can transmit packets directly when needed, without any signaling exchange in the channel, also known as 1-step RACH. Due to this feature, it's named as unsourced multiple access.

An essential presumption of UMA is that only a relatively small group of users turn active simultaneously. The total number of users K can go infinite, while the number of active users K_a always keeps limited, i.e., the active proportion $\mu = K_a/K$ remains nearly constant. The minimum active duration can be set as the length of transmission frame n, in which these K_a users collide. That is to say, it is time-division on a large scale, and code division on a small scale or μ is the mean of Poisson arriving stream per access frame at the BS side. Considering these, K_a generally ranges from 50 to 300.

Due to the short packet traffic feature and limited active proportion assumption, the frame length n is constrained in the range from 30,000 to 120,000. A certain amount of tolerated errors can keep the rate of each user sub-channel from going to zero, according to finite blocklength coding theorem [38]. Inspired by this, Polyanskiy figured out the upper bound of PUPE. The numerical theoretical performance under this framework in [6] reveals that traditional NOMA fails after active users increase to more than 150, causing the rapid increase of SNR required by PUPE. The SNR required by UMA in 1-step RACH is lower than that required by OMA in full RACH to keep the same PUPE in higher user cases. UMA can support massive users and maintain lower access delay and power consumption than CMA and grant-free access, which is more competitive for 6G MTC.

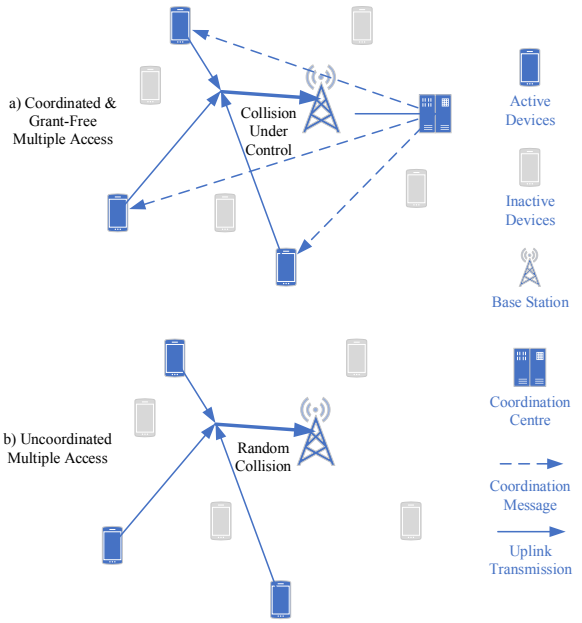


Figure 2. *Transmission behavior comparison among multiple access models for MTC.*

III. CODING SCHEMES OF UNSOURCED MULTIPLE ACCESS

3.1 Designing Paradigms

Although in [6] Polyanskiy presented the theoretical performance bound of UMA, the design of a concrete realizable coding scheme is still an open problem. In general, the overall problem of K_a unsourced users each transmitting k bits on shared n orthogonal resources can be broken down into the following three main factors: 1. Cancellation of multi-user interference; 2. Blind detection under common codebook for unsourced features; 3. Random transmission structure design associated with parameter optimization. For factor 1, a superposition-based multi-access code is required to resolve the collision in the Gaussian Multiple Access Channel (GMAC). There are plenty of solutions in traditional coordinated NOMA coding regimes, some of which can be extended to the uncoordinated scenario, but with necessary modifications for blind detection, as a matter of factor 2. Moreover, the basic multi-access code is enhanced for the concern of performance or complexity. In UMA, the collision pattern of coded packets is not deterministic. However, specific coding structures such as uncoordinated random CDMA [39] or time-division slotted ALOHA [40] can determine the distribution characteristics of packet collisions, which is related to their coding parameters. For factor 3, it is vital to associate the overall optimization of PUPE with the adjustment of specific parameters under these transmission structures to guide the deployment of these coding schemes. Therefore, we will summarize the various UMA coding schemes in these three aspects in the following subsections.

3.1.1 Basic Superposition Multi-Access Code

Collision-resolving can be modeled as the problem of solving linear equations. BCH MA code is designed in $GF(2^k)$ under this regime [41]. According to the arithmetic character of its parity check matrix, the linear combination of no more than T columns can always be de-mapped to the selected components. Thus, each user's k information bits are mapped to the 2^k-1 column vectors of length KT with the rate $1/T$. The result of superposition can be regarded as a syndrome, so the decoder is the same as BCH Forward Error Correction (FEC) code, using Berlekamp-Massey algorithm [42] to iteratively resolve the exponentials of error polynomial, i.e., the column index of each syndrome component.

Compressed sensing MA code models this problem in the real or complex field. For any T-sparse random matrix satisfying restricted isometry property (RIP) constraint $N_s > CT\ln(2^k/T)$ (C is some constant), the superposition of no more than T columns are resolvable [43]. The decoding of CS MA code is the problem of support set recovery. Greedy algorithm Orthogonal Matching Pursuit (OMP), minimization method l_1-LASSO (Least Absolute Shrinkage and Selection Operator) with threshold, and message passing based method AMP (Approximate Message Passing) are effective choices [44]. Moreover, the CS MA code directly maps the k information bits to N_s symbols. Due to the spreading gain, there is no need for an extra FEC code.

From the superposition coding perspective, there are other pathways. Spatially coupled LDPC (SC-LDPC) [45] encodes each user bits with a common original LDPC (Low Density Parity Check) and then permutes the codewords with different patterns before superposition. Moreover, IDMA [46] is the more generalized interleaver-based coding scheme, consisting of a low rate FEC (usually generated by repetition of an original FEC code) and a user-determining interleaver. Since the superposition of permuted codewords can be transformed as the extension of the factor graph, both of them can be decoded with turbo-like iterative Soft-Input-Soft-Output (SISO) structure [47] or multi-layer Belief Propagation (BP) on extended factor graph[48].

Polar MA code is designed based on list decoding with Successive Interference Cancellation (SIC) structure [49]. Information bits of each user is appended with Cyclic Redundancy Check (CRC) bits before polar encoding to enable the Tal-Vardy's Successive Cancellation List (SCL) decoding [50, 51] at the receiver end. Due to the polarization effect, different codewords of polar code have different transmission qualities. The interference from other users is treated as noise (TIN), so the superposition of polar codewords can be decoded successively. The output of the list decoder that passes the CRC check is recovered as the information bits of one of the users, which is then remapped for SIC. Thus, all the information from different users can be recovered successively under this TIN-SIC iteration [52].

3.1.2 Enhancement and Modification for Unsourced Scenario

Most of the above mentioned MA codes work successfully in traditional MA modes, while various degrees of modification or enhancement are required for them to adapt to the uncoordinated GMAC scenario. Although in theory Bose–Chaudhuri–Hocquenghem (BCH) MA code is capable of resolving collisions, the arithmetic character of the transmitted syndrome can be easily corrupted by noise. All of the decoding results would be wrong if any bit of the syndrome changes. To ensure the transmission under AWGN, a linear FEC is successively concatenated to it. BCH MA code is utilized as the basic outer code with a inner FEC in Concatenated Code (CC) scheme [7], which is the first UMA coding scheme proposed by Polyanskiy and Ordenlitch.

While for CS MA codes, the main problem lies in complexity. Under the basic setup for $k = 100$, the column size of sensing matrix should be at least 2^{100}, causing unacceptable storage and decoding complexity costs. Coded Compressed Sensing (CCS) [53] adopts the spirit of divide and conquer, splitting the chunk of information bits into sub-blocks, each encoded with the downsized sensing matrix. Meanwhile, to stitch these sub-blocks together for every user separately, additional tree code [54] bits is added to each sub-block to introduce successive parity check relationship. The tree decoder is appended to per-block CS decoder to validate the association between sub-blocks. Moreover, to further reduce the decoding complexity of CS MA code, a structured sensing matrix design with according enhanced AMP algorithm is

proposed under CCS scheme by [55].

IDMA requires user-specific interleavers to separate different user-palyloads, which can't be preconfigured by coordination center in UMA. To preserve common codebook feature, the interleaver pattern of IDMA in UMA is determined by part of the user information bits for randomization and efficiency. The transmitted packet should contain two parts, the IDMA codeword and the according coded pilot that usually mapped by CS code. The decoder first detect the coded pilot to complete the expansion of superpositioned factor graph and then start the iterative decoding process. Eventually the IDMA decoding results of pilot and IDMA codeword are stitched together at the end of each decoding branch individually. [56] designed a CS pilot based coding structure for SC-LDPC to identify users, enabling the BP and SIC procedure at the receiver end. Besides, in sparse IDMA scheme [57] the pilot part determines not only the interleaver pattern, but also the repetition rate of each user's FEC, introducing irregular superpostion pattern and reducing collision among symbols with zero-padding.

The leveraging of polar MA codes is mainly based on spreading techniques. Since the decoding process begins in TIN mode, appending random [58] or sparse spreading [59] after polar MA encoder can bring coding gain and reduce the interference. Meanwhile, the spreading sequence can be detected and equalized during SIC process to help the start of list decoding.

3.1.3 Transmission Structure Design

From the universal view, the transmission structure of UMA can be modeled as a GMAC of K_a users with k-bit input and n-bit codeword each by superposition, depicted as figure 3-c). This works for the design of polar random spreading [58] and sparse spreading [59] schemes, as long as the sparse IDMA [57] scheme taking nearly all n bits to construct an ultra low rate code with only one pilot, while it is not suitable for most of the remaining schemes. For instance, Designing a CCS code directly of 100 info-length and 30,000 code-length would be difficult and ineffective. Except for the direct spreading structure, there are two important intermediate ones, the T-Fold ALOHA [7] and T-Fold RSA [60] (Repetition Slotted ALOHA).

In T-Fold ALOHA, the total n resources are divided into V slots of length $\bar{n} = n/V$ each. K_a users randomly choose one and only one slot to transmit coded packets, which is uniformly distributed, as shown in figure 3-a). The number of collided packets on each slot follows a binomial distribution with K_a times of experiments and probability $1/V$ for each, i.e., $L \sim \mathcal{B}(K_a, 1/V)$. Thus, "T-Fold" means to control the probability of the number of users superimposed on any slot L under the collision threshold T. In other words, the required capability of collision resolving of MA codes under T-Fold ALOHA is at most T. Collision L is positively correlated with the number of users K_a and negatively with the number of slots V. To ensure $P(L > T) \to 0$, it's necessary to reduce $\bar{n}$ when K_a increases. With the same code-length $\bar{n}$, as analyzed before, the SNR requirement of MA codes increases with the increase of T. Therefore, the trade-off among these coding parameters is the optimization problem for bound approaching under the T-Fold ALOHA structure. Most single slot MA schemes such as BCH-based CC [7], IDMA based schemes [57], SPARC [61, 62] adopt this structure.

As for T-Fold RSA, part or all users repeat their packets for certain or random times, introducing diversity gain through packet coding and enabling packet-level SIC at the receiver. This is proposed in a series of Liva's articles [63–65] studying the optimization of packet degree distribution for random access, proving that the irregular repeated packet coding IRSA (Irregular Repetition Slotted ALOHA) outperforms the regular CRDSA (Contention Resolution Diversity Slotted ALOHA). As shown in figure 3-b), given the transmitting degree polynomial $\lambda(x)$, $L \sim \mathcal{B}(\lambda'(1)K_a, 1/V)$, i.e. the distribution of L is controlled by $\lambda(x)$. Like T-Fold ALOHA, this multi-slot coding scheme has a collision threshold of T for MA code within one slot. However, for the cases of $L > T$, the chances are that there exist successfully decoded replicas on other slots that $L \leq T$. After SIC, some threshold exceeding cases can be reduced under T for MA decoding. This process is similar to the decoding of LDPC under BEC(Binary Erasure Channel), and the analysis tool can also be applied to the optimization of $\lambda(x)$. In [66], the degree of repetition and T-Fold MA code are optimized by modified DE iteration and FBL Bound. Especially for UMA, the number of packet repetitions is given by a local random number generator following optimal λ for each user. The parameter optimization of the CCS scheme with splitting [53] can be re-

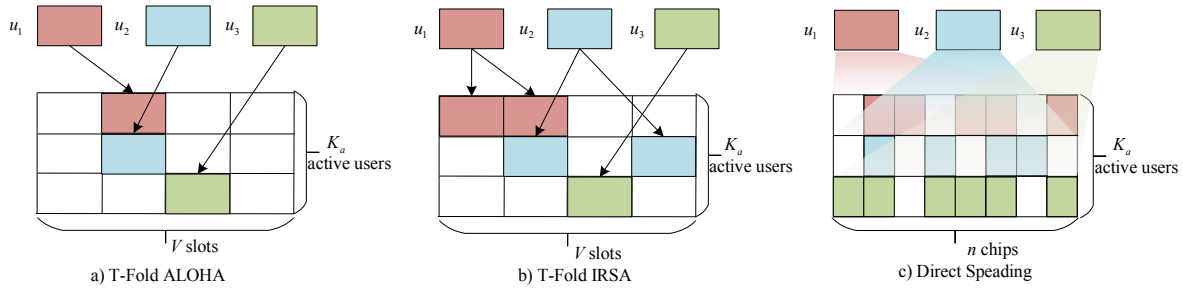


Figure 3. *Transmission structures of unsourced multiple access.*

garded as T-Fold CRDSA in terms of parameter optimization, although the multiple packets are not replicas. SC-LDPC with SIC [56] and Polar-IRSA [67] are designed under T-Fold IRSA.

To conclude, the existing coding schemes can be categorized by the above three determining paradigms. Table 1 gives a summary of mentionable solutions in published works.

3.2 State-of-Art Coding Schemes

In this subsection, the detailed transmitter and receiver structures of three selected typical coding schemes of UMA in table 1 are introduced. Sparse IDMA utilizes the enhanced IDMA as MA code and T-Fold ALOHA as transmission structure. CCS uses CS for MA code with tree code for splitting and stitching under the T-Fold CRDSA. Polar spreading combines polar MA code with direct spreading. All of them are two-level concatenated code in some sense and representative in their respective categories. Without loss of generality, the sketches in figure 4-6 of each are simplified as two-user GMAC.

The system structure of sparse IDMA [57] is presented in figure 4. As mentioned in 3.1, sparse IDMA is designed based on IDMA with pilot, adding random irregular repetition within IDMA superposition coding sctructure. The information bits $\boldsymbol{w}_i$ from user i of length k is splitted into two parts, $\boldsymbol{w}_i^s$ with length k_s and $\boldsymbol{w}_i^c$ with length k_c. $\boldsymbol{w}_i^s$ is coded by $N_s \times 2^{k_s}$ CS matrix for pilot $\boldsymbol{x}_i^s$, also as the seed of pseudo-random number generator to determine number of repetition following certain degree distribution $\alpha_i(\boldsymbol{w}_i^c) \sim \lambda(x)$ and the index of interleaver pattern $f_i \in \mathcal{F}$. $\boldsymbol{w}_i^c$ is first sent to LDPC encoder with rate RL, the BPSK modulated codeword of which is denoted by $\boldsymbol{u}_i$. And $\boldsymbol{u}_i$ is then repeated α_i times and zero-padded to the length of N_c, to generate sparsely spread codeword $\boldsymbol{u}_i^r$. The BPSK modulated symbol sequence $\boldsymbol{u}_i^r$ is interleaved by $\boldsymbol{x}_i^c = f_i(\boldsymbol{u}_i^r)$. Finally $\boldsymbol{x}_i^s$ and $\boldsymbol{x}_i^c$ are stitched together to form the coded symbol packet $\boldsymbol{x}_i$ of one slot length $\bar{n} = N_s + N_c$. For $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ colliding in one slot, their pilot parts $\boldsymbol{x}_1^s$ and $\boldsymbol{x}_2^s$ and IDMA codedword parts $\boldsymbol{x}_1^c$ and $\boldsymbol{x}_2^c$ collide in the channel individually. Under this T-Fold ALOHA structure, the row and column size of CS pilot encoder and IDMA coding setup RL together with $\lambda(x)$ are the main parameters to be optimized to guarantee the "T-Fold" condition. The encoding process is performed in parallel, while the decoding is serial. The factor graph of superimposed LDPC code is reconstructed by recovering the support set of pilots $\hat{\boldsymbol{w}}_i^s$ from $\boldsymbol{y}^s = \sum_{i=1}^{L} \boldsymbol{x}_i^s$ using algorithms like OMP. Then through BP iteration on this 2-layer L-branch irregular factor graph, the IDMA information bits $\hat{\boldsymbol{w}}_i^c$ is recovered at variable nodes of each branch [48]. Due to the bijective relationship between branches and patterns, the decoded $\hat{\boldsymbol{w}}_i^s$ and $\hat{\boldsymbol{w}}_i^c$ are stitched again to form the final result $\hat{\boldsymbol{w}}_i$ of detection.

Unlike sparse IDMA, CCS [53] is a multi-packet coding scheme. $\boldsymbol{w}_i$ is splitted into J sub-blocks $\boldsymbol{w}_i^{(j)}$ with index j and length $k^{(j)}$ for tree coding. Except the first heading sub-block $\boldsymbol{w}_i^{(1)}$, all the following sub-blocks are appended with $c_i^{(j)}$-bit tree code parity check $\boldsymbol{t}_i^{(j)}$ to output coded sub-blocks $\boldsymbol{u}_i^{(j)}$ with equal length $G = k_i^{(j)} + t_i^{(j)}$. Then, each tree coded sub-block $\boldsymbol{u}_i^{(j)}$ is sent to the same $N_s \times 2^G$ CS encoder to generate spreading codeword $\boldsymbol{x}_{i,v}^{(j)}$, transmitted on consecutively randomly chosen slot v. Therefore, each user encoder has J parallel coded packets, utilizing the T-CRDSA transmission structure with J diversity.

Table 1. *The existing UMA coding schemes.*

Coding scheme	Basic MA code	Modification	Transmission structure
CC [7]	BCH MA code	concatenated with FEC	T-Fold ALOHA
CCS [53]	Compressed sensing	Tree code based slicing and stitching[54]; AMP [68], e-AMP [55] enhanced decoding for CS code	T-Fold ALOHA/ T-Fold CRDSA
SPARC [61]	Sparse regression code	AMP CS decoding [62]	T-Fold ALOHA
SC-LDPC+SIC [56]	SC-LDPC	CS pilot header	T-Fold IRSA
Sparse IDMA [57]	IDMA	CS pilot header	T-Fold ALOHA/ Direct spreading
Polar-IRSA [67]	Polar MA code	SCL decoding; packet repetition	T-Fold IRSA
Polar-SS [59]	Polar MA code	Sparse spreading	Direct spreading
Polar-RS [58]	Polar MA code	Random spreading	Direct spreading

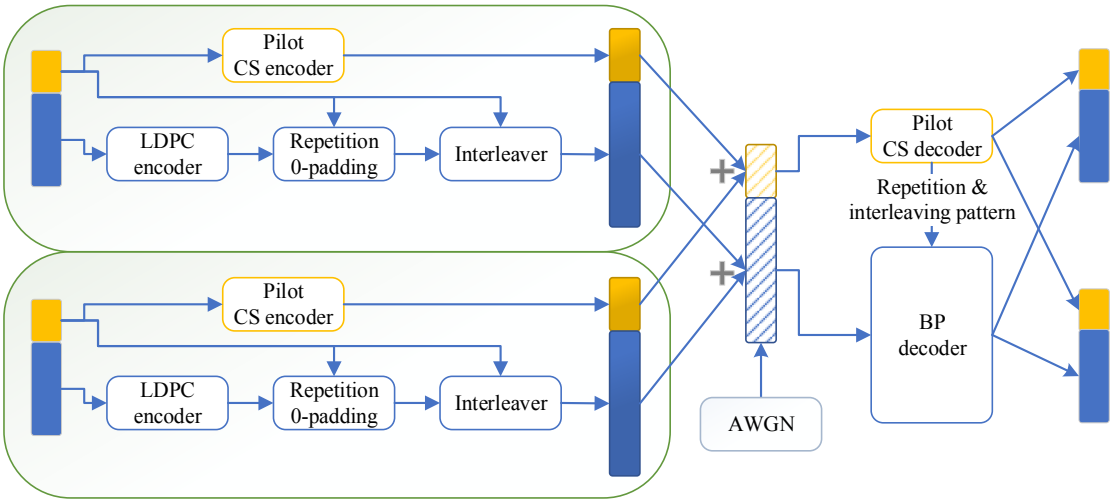


Figure 4. *System structure of sparse IDMA.*

The CS encoder has to ensure T-sparse resolvability and the tree encoder is designed to stitch J sub-blocks together. The decoder is the mirrored reverse of the encoder. With CS decoding slot-by-slot, the L_v superimposed packets $\boldsymbol{y}_v$ on each slot v are first recovered as support set and then remapped as sub-blocks $\hat{\boldsymbol{u}}_{i,v}^{(j)}$ of tree codeword. Then from the first uncoded anchor sub-block $\hat{\boldsymbol{u}}_{i,v}^{(1)}$ of each user i, the parity check relationship between the neighbor slots are validated. One survive path is selected from L_v paths on each stage. Thus, the decoding process for each user is to find a surviving path on a depth-J tree with depth-first style. These $J \times K_a$ sub-blocks are finally threaded as K_a recovered information chunks $\hat{\boldsymbol{w}}_i$ of each user.

The coding structure of the polar spreading scheme [58] features a two-stage bit-level spreading style. Similar to the pilot encoding of sparse IDMA, the spreading sequence $\boldsymbol{s}_i$ is determined by $\boldsymbol{w}_i^s$, a part of information bit sequence $\boldsymbol{w}_i$. Through hash mapping $g(\cdot)$, $\boldsymbol{s}_i^s = \boldsymbol{d}_j$ is selected from the sequence set $\mathcal{D}$ by the index $j = g(\boldsymbol{w}_i^s)$. However, the polar coded and modulated codeword $\boldsymbol{u}_i$ is not transmitted in parallel with $\boldsymbol{s}_i$ as in sparse IDMA, but directly spread by it. The Kronecker product $\boldsymbol{x}_i = \boldsymbol{u}_i \otimes \boldsymbol{s}_i$ from K_a users are superimposed in the channel to generate receiving signal $\boldsymbol{y}$. Under direct spreading structure, the n degrees of freedom of the channel are used as one codeword. The aim of spreading sequence design is minimizing user interference and maximizing spreading gain, with random spreading for pursuit of performance and sparse spreading reducing the complexity. At the receiver, the decoding process is performed in an iterative SIC style. At the start of the t-th iteration, with blind sequence detection, the subset $\hat{\mathcal{D}}^t$ of the set of sequences $\mathcal{D}$ used by all the K_a users is identified. In [58], a segmented energy detector for ran-

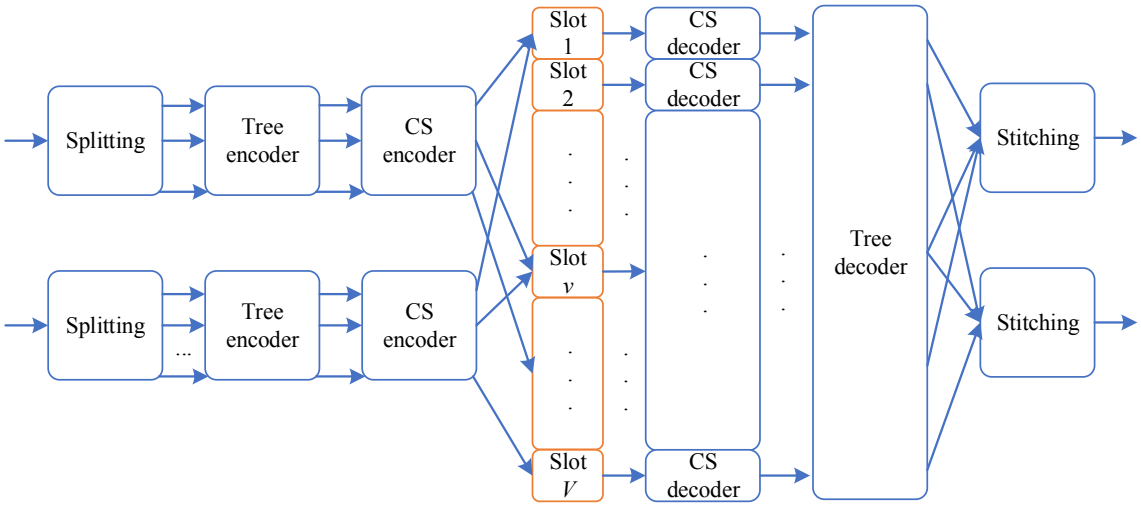


Figure 5. *System structure of CCS.*

dom spreading is introduced. Then an Minimum Mean Square Equalizer (MMSE) [69] based on $\hat{\mathcal{D}}^t$ is performed to produce de-spread soft symbol sequences $\hat{\boldsymbol{\beta}}_i^t$ as in random CDMA. Under polar SCL decoding, each decoding result $\hat{\boldsymbol{w}}_i^{c,t}$ of $\hat{\boldsymbol{\beta}}_i^t$ is checked by CRC to verify its correctness [70]. The set of user index that passes the validation is denoted as $\mathcal{E}$. The decoded sequences $\hat{\boldsymbol{w}}_i^{c,t}$ with $i \in \mathcal{E}$ are remapped to code chips $\hat{\boldsymbol{x}}_i^c$ and then subtracted from the receiving signal $\boldsymbol{y}$. The input of next iteration is $\boldsymbol{y}^{t+1} = \boldsymbol{y}^t - \sum_{i\in\mathcal{E}} \hat{\boldsymbol{x}}_i^c$. During each iteration, there are new candidates adding to the successively decoded set $\mathcal{E}$. Until $|\mathcal{E}| = K_a$, the iteration is terminated, i.e. all the user information $\hat{\boldsymbol{w}}_i$ is decoded.

3.3 Comparison and Evaluation of UMA Coding Schemes

This section gives some analytical and evaluative reviews among the existing coding schemes after introducing them. We focus on two main aspects, the energy efficiency performance and corresponding computing complexity at the receiver end under the same setup of packet traffic loads and error requirements.

3.3.1 Energy Efficiency Performance Evaluation

As the analytic system established by Polyanskiy in his achievability bound research [6], the general evaluation of the performance of UMA coding schemes is the SNR threshold vs. number of users K_a curve, which is the reflection of system energy efficiency. The SNR threshold is the minimum required E_b/N_0 to keep the PUPE of the user payloads below a certain level ϵ, under the condition that the number of active users K_a and the total degrees of freedom of the channel n is limited. With given n, k, and ϵ, the SNR threshold rises to counter the strengthening of user interference brought by more active users K_a. The bound of SNR threshold under random coding as the best achievable performance is given in [6] as depicted in figure 7, which is the approaching target of UMA coding scheme design.

As summarized in Table 1, the compared schemes can be categorized by their transmission structure in three groups. Under the T-Fold ALOHA structure, the BCH MA code-based CC scheme [7] is worst in performance. It is mainly because of the vulnerability to errors of BCH MA code, relying on the BLER of the short FEC (Forward Error Correction) inner code. Moreover, the maintenance of "T-Fold" when K_a increase is realized by the rapid rising of BCH code length. Thus, the curve has a large gap of 10 dB from the bound and a sharp tangent in the growth trend.

As for CS-based schemes, the CCS-SIC [53] has significant improvement against the CC scheme. Compressed sensing directly maps user information to the complex signal domain and handles the user interference by resolving support set recovery under noise,

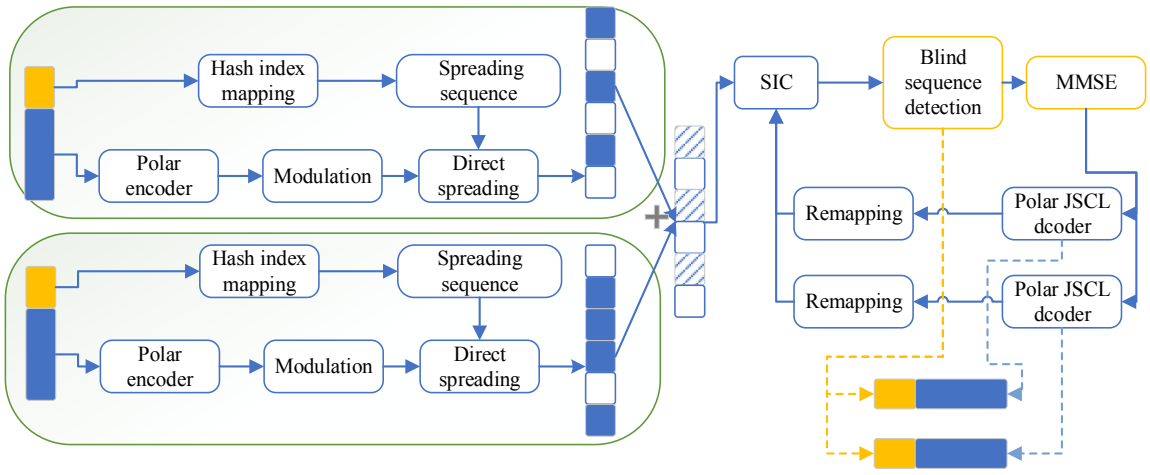


Figure 6. *System structure of polar-RS.*

which improves system energy efficiency. Meanwhile, the tree code is also effective enough to enable downsized CS coding and sub-block stitching without much loss of code rate. The CCS-eAMP [55] adjusts the serial structure of CS AMP decoding and tree decoding, inserts the tree parity check into the CS decoding process between adjacent slots, enables pre-pruning operation on the tree structure, and reduces the range of support set to improve the decoding efficiency. Therefore, this enhanced decoding structure has almost 2dB gain compared with the standard CCS-SIC scheme under lower K_a cases. However, The overall SNR threshold curves of the CCS systems increase rapidly. In order to ensure the sparsity condition, the number of collisions L_v on a single slot needs to be controlled. Therefore, when V increases and $\bar{n}$ becomes shorter, the CS performance deteriorates rapidly. The depth and breadth of the code tree expand simultaneously, causing higher error rates while stitching.

The sparse IDMA [57] has the best performance among all T-Fold ALOHA schemes. Due to the irregular spreading of low rate FEC bits within each packet, the BP iteration effectively canceled the user interference on the optimized superposition factor graph structure. The relatively slow increase of SNR is introduced by the reduced degrading effect of sparse IDMA. However, it still has 5dB gap under $K_a = 200$ setup. Since the collisions increase rapidly with K_a, to keep the "T-Fold" feature, V needs to increase while $\bar{n}$ to decrease. The gain caused by irregular repetition gradually weakens with shortened code length.

Compared with T-Fold ALOHA schemes, the T-Fold IRSA design relaxes the requirement for collision tolerance T is relaxed through package-level SIC enabled by optimized transmission diversity gain $\lambda(x)$ under the same K_a, to lower the SNR threshold. Within this category, the main differences between the two IRSA schemes, the SC-LDPC+SIC [56] and polar-IRSA [67], are the choice of basic MA code and gain of coding parameter optimization. Polar MA code under SC list decoding has better collision tolerance T than SC-LDPC. The curve of SC-LDPC+SIC is attained under $T = 4$ while the polar-IRSA is under $T = 8$. Compared with sparse IDMA in the T-fold ALOHA model, this scheme is still limited in its ability to counter user growth, indicating that the joint optimization between basic MA code and the irregular packet diversity needs to be further enhanced.

Finally, the two most effective coding schemes are polar-random spreading (polar-RS [58]) and polar-sparse spreading (polar-SS [59]) utilizing direct spreading technique. The SNR threshold of polar-RS is very close to the bound under low K_a cases and most close over the existing schemes. As a complexity-reduced version of polar-RS, the polar-SS maintains the performance advantage with slight degradation. All the performance curves of these above mentioned are synthetically compared in figure 7.

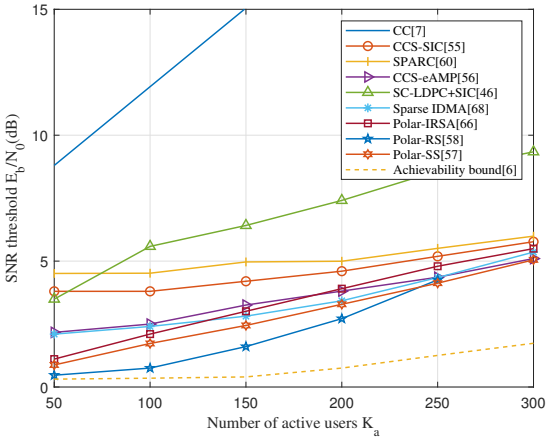


Figure 7. K_a *vs* E_b/N_0 *curves at PUPE=0.1 for the setup* $k = 100$ *bits and* $n = 30,000$.

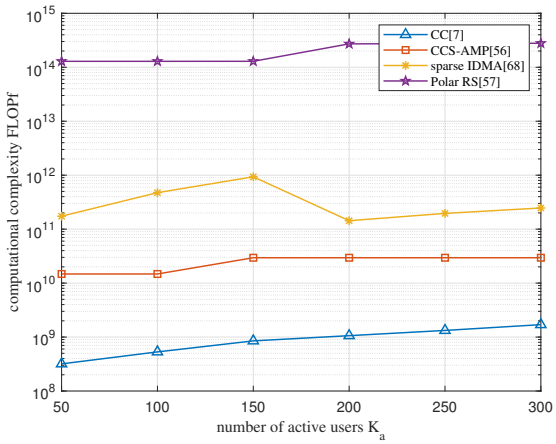


Figure 8. K_a *vs. FLOPf curves at PUPE=0.1 for the setup* $k = 100$ *bits and* $n = 30,000$.

3.3.2 Computational Complexity Analysis

Due to the unsourced uplink feature of UMA, the complexity of the whole system mainly concentrates on the decoding process of the receiver. From the perspective of the coding paradigm, each factor in subsection 3.1 can affect the complexity of the decoding algorithm. Direct spreading and T-Fold IRSA transmission structure need to receive the full length-n frame and perform the overall packet decoding or chip decoding SIC process. However, the T-Fold ALOHA mode can be used for independent slot decoding and has more significant advantages in a processing delay. The choice of basic MA code determines the complexity of single packet decoding. BCH with Berlekamp-Massey algorithm performs decoding on $GF(2^k)$ and has the lowest complexity. The AMP decoding algorithm contributes the most complexity of CS-based schemes. The factor graph size increases with the size parameters of sensing matrix N_s and 2^G. As for tree coding, its complexity is about the same as the depth-J first traversal process on the sub-block parity check tree. The decoding of sparse IDMA consists of two stages, the AMP decoding of the pilot and BP decoding on the expanded factor graph, both of which are iterative message-passing algorithms. The energy detector and MMSE expenses the most computational cost of the polar-RS scheme because of the blind detection on the large dimension of random spreading.

For the above four mentioned typical coding schemes, table 2 gives a detailed computational complexity as the function of coding parameters. Note that the measure of computational complexity, the FLOPf (floating-point operations per frame), is based on the expected total floating-point calculation of decoding all the user packets in one access frame of length n. This definition and granularity of complexity guarantee fairness when it comes to different coding schemes. Moreover, under this framework, we substitute the optimized parameters in [7, 68, 57, 58] to get the K_a vs. FLOPf curves of each coding scheme. For these encoding schemes in figure 8, the primary trend in computational complexity is to increase as the number of users increases, but there are vast orders of magnitude differences among them. Since their SNR threshold varies widely, it is necessary to compare the complexity cost of energy efficiency gain or trade-off relationship between these two evaluation merits among different coding schemes.

In figure 9, the number of users is fixed at medium load setup $K_a = 150$. Comparing CC and CCS-AMP, it is clear that the SNR threshold is reduced by about 10dB while the complexity increases by only one order of magnitude. However, going from CCS-AMP to Sparse IDMA costs more than an order of magnitude of complexity for an efficiency gain of only 1.4dB. The 1.2dB performance improvement from Sparse IDMA to Polar- RS requires another two orders of magnitude of complexity. This phenomenon indicates that the closer it gets to the bound, the more complexity it needs to pay for marginal leverage of the performance. The comprehensive analysis of these existing coding schemes shows that some are energy-efficient enough

Table 2. *Computational complexity comparison of UMA coding schemes.*

Coding scheme	Related parameters	Computational complexity
CC [7]	k dimension of GF; T threshold of collision; Z_c LDPC lifting size; $I_{\max}$ maximum iteration of SPA; n_v number of variable nodes; n_c number of parity check nodes; V number of slots	$V[2m(T+m)+$ $\left(3Z_c\left(m_c+n_v\right)^2 I_{\max}\right)]$
CCS-AMP [68]	N_s row size of CS matrix; 2^G column size of CS matrix; K_a number of users; J number of sub-blocks; $I_{\max}$ maximum iteration of AMP	$4JI_{\max}K_aN_s\left(2^G+1\right)+K_a\log_2\left(K_a\right)$
Sparse IDMA [57]	N_p row size of CS matrix for pilot coding; 2^G column size of CS matrix for pilot coding; K_a number of users; $RR=\lambda'(1)$ average repetition rate; RL basic LDPC code rate; $N_c=n-N_p$ IDMA code length; $N_c=k-B_p$ IDMA information bit length; I_1 BP user layer iteration; I_2 BP channel layer iteration	$4N_sK_a\left(2^G+1\right)+$ $I_2\left[I_1K_aN_c^2((RR+1)/RL-1)^2$ $+(K_a*RR*N_c/RL+n)^2\right]*3$
Polar-RS [58]	2^{B_s} total number of spreading sequences; N_s length of spreading sequences; N_c length of information bits for polar coding; r length of CRC parity check bits; B_c polar code length; g length of segment in energy detector; $I_{\max}$ maximum iteration of SCL decoder	$2n\left(2^{B_s}\right)2^gB_c/g+2N_s2^{B_s}+$ $K_a\left(N_s^22^{B_s}+2^{(3B_s)}+2N_s2^{B_s}+\right.$ $\left.I_{\max}B_c\log 2\left(B_c\right)+B_cr\right)$

to approach the ideal bound at low user cases. However, the complexity cost of these solutions is too high, so a better trade-off between complexity and performance is needed. Meanwhile, there is still a big gap between the energy efficiency and the performance of the existing solutions under high user numbers, and the design needs to be further optimized to cancel the fierce multi-user interference.

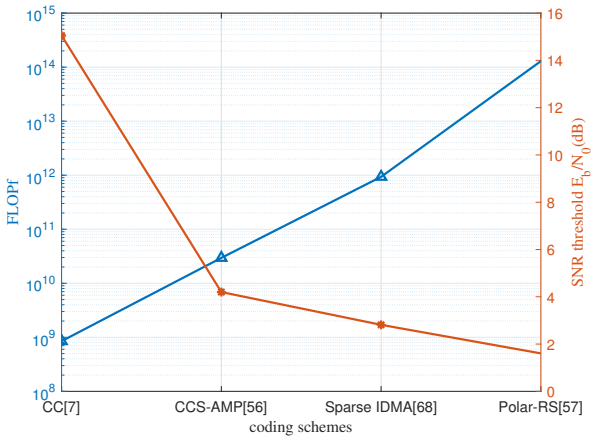


Figure 9. *Performance and computational complexity comparison at $K_a = 150$, PUPE=0.1 for the setup $k = 100$ bits and $n = 30,000$.*

IV. OPEN ISSUES AND RESEARCH POTENTIALS

4.1 Bound Approaching Coding Scheme Design

In terms of performance, lowering the SNR threshold is the permanent focal point of UMA code design, which is vital for the low-power feature of devices in MTC Networks. In addition to the existing schemes adapted from traditional with modification in various degrees. Besides, although the performance curves of some existing schemes are pretty close to the achievability bound at low load cases, they diverge from the bound quickly when K_a becomes higher. The optimization of code design to narrow the gap, especially in the high K_a region, remains to be done. The following competitive coding scheme is expected to tackle the increase of multi-user interference when K_a grows to have a better tangent at the K_a-SNR curve.

4.2 Complexity Reduction

According to the analysis in section 3.3.2, the cost of complexity for performance improvement increases rapidly. Although the computational resources are assumed to be abundant at the receiver side, the complexity problem is always a genuine concern for manufactures and operators when it comes to deployment. The access points (AP) in MTC network should be low-cost and simple for massive deployment. As the analysis shows, the complexity problem relies on decoding algorithms and the coding scheme's overall design. Existing schemes have already considered the realization complexity in their design, but this is not enough. The trade-off between performance and complexity is another essential work excepted from the competition of extreme performance.

4.3 Practical Propagation Conditions: Fading & Asynchronous

The mainstream research results in this field focus on the coding design under the GAMC channel, while few consider the implementation under the more realistic propagation models. The problems that we must tackle are practically the block fading characteristics on different user packets and the synchronization between blocks caused by unscheduled features, which is typical for MTC scenarios. [71] gives the approximate performance bound for fading channels and an ideal coding and decoding framework. [72] preliminarily considers the influence of block misalignments in the case of asynchronous and fading channels. The coding scheme design with performance advantages in asynchronous and fading channels is also an important work to consider for promoting the application and standardization process of UMA for MTC. Blind equalization, channel estimation, and the corresponding coding structures are the main techniques to be developed.

4.4 MIMO UMA

Massive MIMO has become the key technology in the 5G system and will be evolving to 6G. MIMO can introduce more diversity gain by multiple antennas, and joint space-time code design can increase the degree of freedom in the system. At present, [73, 74] combines UMA coding scheme with MIMO, but their optimization for energy efficiency and MIMO fading channels are still insufficient. The capacity gain of MIMO has been widely studied, but its effect on the PUPE under UMA is not clarified. With distributed MIMO [75] deployed in uplink transmission, the design of two-dimensional time-space transmission structure and the maximization of multi-antenna diversity coding gain are two crucial research focuses. The existing results of blind detection [76], multi-user MIMO design [77], and pilot coding design [78] in the MIMO field can provide more research ideas for the design of UMA.

4.5 Physical Layer Security

The 1-step access process of the UMA system does not identify and authenticate user data packets under MTC scenarios. In some MTC network operations, it is necessary to check the identity and behavior legitimacy of access devices, e.g. security sensor data transmission. The total openess of air-interface may encounter malicious attacks. In addition, the passive data packet may have eavesdropped because of the openness of the air interface and the universality of the codebook. Therefore, it is necessary to introduce some low-complexity physical layer security [79] technologies to the code of the UMA on the sending end and add the packet decryption process on the receiving end to ensure the secure and effective transmission of short packets.

ACKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China under Grant 61971062, Grant 92067202, Grant 62071058, Grant 62001049, Beijing Natural Science Foundation under Grant 4222012, and Beijing University of Posts and Telecommunications - China Mobile Research Institute Joint Innovation Center.

References

[1] 3GPP, "Release 16 description; summary of rel-16 work items," 3GPP, Tech. Rep. 21.906, Jun. 2021.
[2] B. Zong, C. Fan, *et al.*, "6G Technologies: Key Drivers, Core Requirements, System Architectures, and Enabling Technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, 2019, pp. 18–27.
[3] Q. Bi, "Ten Trends in the Cellular Industry and an Outlook on 6G," *IEEE Communications Magazine*, vol. 57, no. 12, 2019, pp. 31–36.

[4] Z. Zhang, Y. Xiao, *et al.*, "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, 2019, pp. 28–41.

[5] Y. Wu, X. Gao, *et al.*, "Massive Access for Future Wireless Communication Systems," *IEEE Wireless Communications*, 2020, pp. 1–9.

[6] Y. Polyanskiy, "A perspective on massive random-access," *IEEE International Symposium on Information Theory - Proceedings*, no. 2, 2017, pp. 2523–2527.

[7] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," *IEEE International Symposium on Information Theory - Proceedings*, 2017, pp. 2528–2532.

[8] F. Boccardi, R. W. Heath, *et al.*, "Five disruptive technology directions for 5g," *IEEE Communications Magazine*, vol. 52, no. 2, 2014, pp. 74–80.

[9] A. Al-Fuqaha, M. Guizani, *et al.*, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, 2015, pp. 2347–2376.

[10] G. A. Akpakwu, B. J. Silva, *et al.*, "A survey on 5g networks for the internet of things: Communication technologies and challenges," *IEEE Access*, vol. 6, 2018, pp. 3619–3647.

[11] A. Hoglund, X. Lin, *et al.*, "Overview of 3gpp release 14 enhanced nb-iot," *IEEE Network*, vol. 31, no. 6, 2017, pp. 16–22.

[12] N. H. Mahmood, S. Böcker, *et al.*, "Machine type communications: key drivers and enablers towards the 6g era," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, 2021, pp. 1–25.

[13] N. H. Mahmood, H. Alves, *et al.*, "Six key features of machine type communication in 6G," in *2nd 6G Wireless Summit 2020: Gain Edge for the 6G Era, 6G SUMMIT 2020*. Institute of Electrical and Electronics Engineers Inc., mar 2020.

[14] J. He, A. Radford, *et al.*, "Cooperative connected autonomous vehicles (CAV): Research, applications and challenges," in *Proceedings - International Conference on Network Protocols, ICNP*, vol. 2019-Octob, 2019.

[15] S. Chen, J. Hu, *et al.*, "A vision of c-v2x: Technologies, field testing, and challenges with chinese development," *IEEE Internet of Things Journal*, vol. 7, no. 5, 2020, pp. 3872–3881.

[16] H. Kagermann, W. Wahlster, *et al.*, "Securing the future of German manufacturing industry: Recommendations for implementing the strategic initiative INDUSTRIE 4.0," *Final Report of the Industrie 4.0 Working Group*, no. April, 2013, pp. 1–84.

[17] M. Giordani, M. Polese, *et al.*, "Toward 6g networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, 2020, pp. 55–61.

[18] V. Jacobson. (2021, Aug.) Medical wearables: market and technology trends. open access. [Online]. Available: https://s3.i-micronews.com/uploads/2019/03/YD19008-Medical-Wearables-2019-Yole-Sample-2.pdf

[19] X. You, C.-X. Wang, *et al.*, "Towards 6g wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, no. 1, 2021, pp. 1–74.

[20] N. H. Mahmood, H. Alves, *et al.*, "Six key enablers for machine type communication in 6g," *CoRR*, vol. abs/1903.05406, 2019. [Online]. Available: http://arxiv.org/abs/1903.05406

[21] Y. Zhao, J. Zhao, *et al.*, "A Survey of 6G Wireless Communications: Emerging Technologies," in *Advances in Intelligent Systems and Computing*, vol. 1363 AISC. Springer Science and Business Media Deutschland GmbH, 2021, pp. 150–170.

[22] G. Durisi, T. Koch, *et al.*, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE*, vol. 104, no. 9, 2016, pp. 1711–1726.

[23] A. Wolf, P. Schulz, *et al.*, "How reliable and capable is multi-connectivity?" *IEEE Transactions on Communications*, vol. 67, no. 2, 2019, pp. 1506–1520.

[24] M. Wollschlaeger, T. Sauter, *et al.*, "The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0," *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, 2017, pp. 17–27.

[25] 5G-ACIA. (2021, Aug.) 5g for connected industries and automation (white paper - second editon). open access. [Online]. Available: https://www.zvei.org/en/press-media/publications/5g-for-connected-industries-and-\
automation-white-paper-second-editon

[26] S. Nahavandi, "Industry 5.0-a human-centric solution," *Sustainability (Switzerland)*, vol. 11, no. 16, 2019.

[27] A. W. Scott and R. Frobenius, *RF Measurements for Cellular Phones and Wireless Data Systems (Scott/RF Measurements) —— Multiple Access Techniques: FDMA, TDMA, AND CDMA*. Wiley-IEEE Press, 2008.

[28] H. Li, G. Ru, *et al.*, "OFDMA capacity analysis in MIMO channels," *IEEE Transactions on Information Theory*, vol. 56, no. 9, 2010, pp. 4438–4446.

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

[30] A. Benjebbour, K. Saito, *et al.*, "Non-orthogonal multiple access (noma): Concept, performance evaluation and experimental trials," in *2015 international conference on wireless networks and mobile communications (WINCOM)*. IEEE, 2015, pp. 1–6.

[31] L. Dai, B. Wang, *et al.*, "A Survey of Non-Orthogonal Multiple Access for 5G," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, 2018, pp. 2294–2323.

[32] Z. Ding, X. Lei, *et al.*, "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, 2017, pp. 2181–2195.

[33] D. T. Wiriaatmadja and K. W. Choi, "Hybrid random access and data transmission protocol for machine-to-machine communications in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, 2015, pp. 33–46.

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

[35] N. DOCOMO, "Uplink Multiple Access Schemes for NR," RAN1 3GPP Meeting85, Tech. Rep. document R1-165174, jun 2016.

[36] M. Mohammadkarimi, M. A. Raza, *et al.*, "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, 2018.

[37] B. Wang, L. Dai, *et al.*, "Compressive sensing based multi-user detection for uplink grant-free non-orthogonal multiple access," in *2015 IEEE 82nd Vehicular Technology Conference, VTC Fall 2015 - Proceedings*, 2016.

[38] Y. Polyanskiy, H. V. Poor, *et al.*, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, 2010, pp. 2307–2359.

[39] J. Choi, "Performance Analysis of 2-Step Random Access with CDMA in Machine-Type Communication," *IEEE Transactions on Communications*, vol. 69, no. 4, 2021, pp. 2387–2397.

[40] T. Polonelli, D. Brunelli, *et al.*, "Slotted aloha on lorawan-design, analysis, and deployment," *Sensors*, vol. 19, no. 4, 2019, p. 838.

[41] I. Bar-David, E. Plotnik, *et al.*, "Forward Collision Resolution A Technique for Random Multiple-Access to rhe Adder Channel," *IEEE TRANSACTIONS ON INFORMAITON THEORY*, vol. 39, no. 5, 1993, pp. 1671–1675.

[42] T. Polonelli, D. Brunelli, *et al.*, "Slotted ALOHA on LoRaWAN-design, analysis, and deployment," *Sensors (Switzerland)*, vol. 19, no. 4, 2019.

[43] M. Leinonen, M. Codreanu, *et al.*, *Compressed Sensing with Applications in Wireless Networks*, 2019.

[44] N. Koep, A. Behboodi, *et al.*, "An Introduction to Compressed Sensing," in *Applied and Numerical Harmonic Analysis*, 2019.

[45] M. Noor-A-Rahim, K. D. Nguyen, *et al.*, "SC-LDPC code design for Gaussian multiple access channel," *IET Communications*, vol. 10, no. 17, 2016.

[46] K. Kusume, G. Bauch, *et al.*, "IDMA vs. CDMA: Analysis and comparison of two multiple access schemes," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, 2012.

[47] X. Xiong, J. Hu, *et al.*, "A fast converging multi-user detection for IDMA based on time-reversal," in *2007 6th International Conference on Information, Communications and Signal Processing, ICICS*, 2007.

[48] S. Suyama, H. Suzuki, *et al.*, "Iterative multiuser detection with soft decision-directed channel estimation for MC-IDMA and performance comparison with chip-interleaved MC-CDMA," *IEICE Transactions on Communications*, vol. E92-B, no. 5, 2009.

[49] J. Dai, K. Niu, *et al.*, "Polar-coded non-orthogonal multiple access," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, 2018, pp. 1374–1389.

[50] I. Tal and A. Vardy, "List Decoding of Polar Codes," *IEEE Transactions on Information Theory*, vol. 61, no. 5, 2015.

[51] K. Niu and K. Chen, "CRC-aided decoding of polar codes," *IEEE Communications Letters*, vol. 16, no. 10, 2012.

[52] E. Abbe and E. Telatar, "Polar codes for the m-user multiple access channel," *IEEE Transactions on Information Theory*, vol. 58, no. 8, 2012.

[53] V. K. Amalladinne, G. S. Member, *et al.*, "A Coded Compressed Sensing Scheme for Unsourced Multiple Access," *IEEE Transactions on Information Theory*, vol. 66, no. 10, 2020, pp. 6509–6533.

[54] R. Calderbank and A. Thompson, "CHIRRUP: A practical algorithm for unsourced multiple access," *Information and Inference*, vol. 9, no. 4, 2020.

[55] V. K. Amalladinne, J. F. Chamberland, *et al.*, "An Enhanced Decoding Algorithm for Coded Compressed Sensing," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2020-May, 2020, pp. 5270–5274.

[56] A. Vem, K. R. Narayanan, *et al.*, "A User-Independent Successive Interference Cancellation Based Coding Scheme for the Unsourced Random Access Gaussian Channel," *IEEE Transactions on Communications*, vol. 67, no. 12, 2019, pp. 8258–8272.

[57] A. K. Pradhan, V. K. Amalladinne, *et al.*, "A joint graph based coding scheme for the unsourced random access gaussian channel," *CoRR*, vol. abs/1906.05410, 2019. [Online]. Available: http://arxiv.org/abs/1906.05410

[58] M. Zheng, Y. Wu, *et al.*, "Polar Coding and Sparse Spreading for Massive Unsourced Random Access," in *IEEE Vehicular Technology Conference*, vol. 2020-Novem, 2020, pp. 8–13.

[59] M. Zheng, Y. Wu, *et al.*, "Polar Coding and Sparse Spreading for Massive Unsourced Random Access," in *IEEE Vehicular Technology Conference*, vol. 2020-Novem, 2020, pp. 0–4.

[60] E. Paolini, G. Liva, *et al.*, "Coded Slotted ALOHA: A Graph-Based Method for Uncoordinated Multiple Access," *IEEE Transactions on Information Theory*, vol. 61, no. 12, 2015, pp. 6815–6832.

[61] A. Fengler, G. Student, *et al.*, "SPARCs for Unsourced Random Access," *IEEE Transactions on Information Theory*, vol. 9448, no. c, 2021, pp. 1–23.

[62] A. Fengler, P. Jung, *et al.*, "SPARCs and AMP for Unsourced Random Access," *IEEE International Symposium on Information Theory - Proceedings*, vol. 2019-July, 2019, pp. 2843–2847.

[63] G. Liva, "A slotted ALOHA scheme based on bipartite graph optimization," *2010 International ITG Conference on Source and Channel Coding, SCC 2010*, 2010, pp. 1–6.

[64] E. Paolini, Č. Stefanović, *et al.*, "Coded random access: Applying codes on graphs to design random access protocols," *IEEE Communications Magazine*, vol. 53, no. 6, 2015, pp. 144–150.

[65] A. I. Graell Amat and G. Liva, "Finite-length analysis of irregular repetition slotted ALOHA in the waterfall region," *IEEE Communications Letters*, vol. 22, no. 5, 2018, pp. 886–889.

[66] A. Glebov, N. Matveev, *et al.*, "Achievability Bounds for T-Fold Irregular Repetition Slotted ALOHA Scheme in the Gaussian MAC," *IEEE Wireless Communications and Networking Conference, WCNC*, vol. 2019-April, no. 18, 2019, pp. 12–17.

[67] E. Marshakov, G. Balitskiy, *et al.*, "A polar code based unsourced random access for the gaussian MAC," *IEEE Vehicular Technology Conference*, vol. 2019-Septe, no. 18, 2019, pp. 1–5.

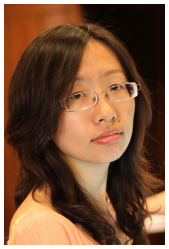
[68] V. K. Amalladinne, A. Department, *et al.*, "On Approximate Message Passing for Unsourced Access with Coded Compressed Sensing," in *IEEE International Symposium on Information Theory - Proceedings*, vol. 2020-June, 2020, pp.

2995–3000.
[69] J. G. Proakis and M. Salehi, "Digital Communications," *Digital Communications*, vol. 73, no. 11, 2015, pp. 247–385.
[70] S. Onay, "Successive cancellation decoding of polar codes for the two-user binary-input MAC," in *IEEE International Symposium on Information Theory - Proceedings*, 2013.
[71] S. S. Kowshik, K. Andreev, *et al.*, "Energy efficient random access for the quasi-static fading MAC," *IEEE International Symposium on Information Theory - Proceedings*, vol. 2019-July, no. 2, 2019, pp. 2768–2772.
[72] K. Andreev, S. S. Kowshik, *et al.*, "Low complexity energy efficient random access scheme for the asynchronous fading MAC," *IEEE Vehicular Technology Conference*, vol. 2019-Septe, 2019.
[73] T. Li, Y. Wu, *et al.*, "SPARC-LDPC Coding for MIMO Massive Unsourced Random Access," *IEEE Globecom Workshops*, 2020, pp. 8–13.
[74] A. Fengler, S. Haghighatshoar, *et al.*, "Non-Bayesian Activity Detection, Large-Scale Fading Coefficient Estimation, and Unsourced Random Access with a Massive MIMO Receiver," *IEEE Transactions on Information Theory*, vol. 67, no. 5, 2021, pp. 2925–2951.
[75] F. Wiffen, M. Z. Bocus, *et al.*, "Distributed MIMO Uplink Capacity under Transform Coding Fronthaul Compression," in *IEEE International Conference on Communications*, vol. 2019-May, 2019.
[76] J. Zhang, X. Yuan, *et al.*, "Blind Signal Detection in Massive MIMO: Exploiting the Channel Sparsity," *IEEE Transactions on Communications*, vol. 66, no. 2, 2018.
[77] J. Zhang, C. K. Wen, *et al.*, "On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, 2013.
[78] S. Ni, J. Zhao, *et al.*, "Optimal pilot design in massive mimo systems based on channel estimation," *IET Communications*, vol. 11, no. 7, 2017, pp. 975–984.
[79] Y. Wu, A. Khisti, *et al.*, "A Survey of Physical Layer Security Techniques for 5G Wireless Networks and Challenges Ahead," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 4, 2018.

Biographies

Yuanjie Li received the B.S. degree in information engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2018. He is currently pursuing the Ph.D. degree with the School of Artificial Intelligence of BUPT. His research interests include information theory, coding theory, multiple access, and signal processing. Email: lyj1996@bupt.edu.cn

Jincheng Dai received the B.S. and Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014 and 2019. He is currently with the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications. His current research interests include information theory, machine learning, and wireless communications. He was a recipient of the China Education Society of Electronics Excellent Doctoral Dissertation Award in 2019, and the Best Paper Award of Chinese Institute of Electronics Information Theory Conference in 2015. He was served as a co-chair for workshop in IEEE WCNC. Email: daijincheng@bupt.edu.cn



Zhongwei Si received the Ph.D. degree from the KTH Royal Institute of Technology, Sweden, in 2013. In 2013, she joined the Beijing University of Posts and Telecommunications, where she is currently an Associate Professor. Her research interests include wireless communication, information theory, and data mining. Email: sizhongwei@bupt.edu.cn

Chao Dong received the B.S. and Ph.D. degrees in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2007 and 2012, respectively. He is currently an Associate Professor with the School of Artificial Intelligence, BUPT. His research interests include MIMO signal processing, multiuser precoding, decision feedback equalizer, and relay signal processing. Email: dongchao@bupt.edu.cn

Jiaru Lin professor and doctoral supervisor at Beijing university of posts and telecommunications (BUPT), gets the bachelor's degree, master's degree and doctor's degree at BUPT, successively. Research and teaching works in the field of intelligent information processing, mainly including: information theory, cognitive radio, channel coding, wireless communication, etc. Email: jrlin@bupt.edu.cn

Sen Wang received the Ph.D. degree from the Beijing University of Posts and Telecommunications. He is currently senior engineer and principal member of technical staff of the Future Research Lab, China Mobile Research Institute. His research interests include optimization with applications to wireless system, 5G/6G air interface technologies, especially on MIMO, waveform and multiple access, radio resource allocation and performance evaluation for future cellular networks. Email: wangsenyjy@chinamobile.com

Yifei Yuan received his Bachelor & Master degrees from Tsinghua University of China, and a Ph. D. from Carnegie Mellon University, USA. He was with Alcatel-Lucent from 2000 to 2008, working on 3G/4G key technologies. From 2008 to 2020, he was with ZTE as technical director and chief engineer responsible for standards research on LTE-Advanced and 5G. Since 2020, he has been with China Mobile Research Institute, responsible for advanced technologies of 6G. His research interests include MIMO, channel coding, non-orthogonal multiple access (NOMA), internet-of-things (IoT), resource scheduling. He has extensive publications, including 6 books on LTE-Advanced and 5G. He is the rapporteur of NOMA study item in 3GPP. He is the recipient of Best Paper Award by IEEE Communications Society Asia-Pacific Board for co-authoring a paper on NOMA in IEEE Communications Magazine. Email: yuanyifei@chinamobile.com