Unsourced Multiple Access: A Coding Paradigm for Massive Random Access

Gianluigi Liva, *Senior Member, IEEE*, and Yury Polyanskiy, *Fellow, IEEE*

***Abstract*—This paper is a tutorial introduction to the field of unsourced multiple access (UMAC) protocols. We first provide a historical survey of the evolution of random access protocols, focusing specifically on the case in which uncoordinated users share a wireless broadcasting medium. Next, we highlight the change of perspective originated by the UMAC model, in which the physical and medium access layer's protocols cooperate, thus reframing random access as a novel coding-theoretic problem. By now, a large variety of UMAC protocols (codes) emerged, necessitating a certain classification that we indeed propose here. Although some random access schemes require a radical change of the physical layer, others can be implemented with minimal changes to existing industry standards. As an example, we discuss a simple modification to the 5GNR Release 16 random access channel that builds on the UMAC theory and that dramatically improves energy efficiency for systems with even moderate number of simultaneous users (e.g., $5-10$ dB gain for $10-50$ users), and also enables handling of high number of users, something completely out of reach of the state-of-the-art.**

***Index Terms*—Random access, multiple access protocols, massive connectivity, channel coding.**

I. Introduction

THE breakneck development of wireless data networks witnessed during the past decades has been continuously driven by new applications. In its early implementations, wireless cellular networks targeted email and messaging services, which required moderate to low data rates. High resolution multimedia content on the Internet and two-way video streaming introduced the need for broadband connectivity. Recently, the rise of massive machine-type communication (mMTC) and Internet of Things (IoT) systems placed a new set of challenges for the design of next-generation wireless systems. These latter use cases entail drastically different features in terms of traffic profile and reliability requirements. As a consequence, new technical solutions that can address the peculiarities of mMTC and IoT systems have been the subject of intense research efforts in recent years. From a medium access point of view, the shift in perspective originating from these new applications is today well understood. Focusing on traffic profiles only, mMTC and IoT systems often foresee large populations of terminals, which are active sporadically and unpredictably and that transmit only small datagrams. This is in stark contrast with the classical setting of broadband connectivity, where the terminal population is typically orders of magnitude smaller than the one of mMTC and IoT systems, and the data exchange between a user and the base station (BS) involves the transmission of large amounts of data, allowing the use of efficient scheduling techniques to handle medium access. Originally, random access protocols (Aloha in [1]) emerged as a technique to enable wireless access connectivity without centralized coordination between users. Although the development of random access techniques predates by several decades the development of cellular networks, they still form a crucial part of modern 5G stacks for the purpose of providing initial access and handling resource requests. Even if perfectly adequate for those use cases in the past, with mMTC/IoT systems random access shall become the main mechanism for transmitting data, thus placing a much stronger emphasis on the need for energy/spectrum efficient protocols and necessitating revamping of the old designs.

The new challenges placed by mMTC/IoT systems led to an information-theoretic treatment of massive random access (MRA) in [2], gave rise to an actively developing field of unsourced multiple access (UMAC), and revived interest in the design of random access schemes. The paradigm shift manifested by the UMAC is conceptually simple: Instead of relegating the details of random access to the medium access control layer, one should leverage additional side information obtained from the physical layer. In doing so, the random access problem can be formulated as a *coding problem* [2]. It is important to emphasize that UMAC provides a foundation to construct powerful MRA schemes for *uncoordinated* uplink channels. This is different from a wide body of work on next generation multiple access (NGMA) protocols for *coordinated* downlink / uplink transmissions, which can be addressed by orthogonal and/or non-orthogonal multiple access (NOMA) schemes. The latter is outside of the scope of this paper as we are exclusively focused on the uncoordinated uplink.

A. Objective and Main Contributions

In this paper, we pursue three objectives, namely:

A. We aim at providing a tutorial introduction to the field of MRA protocols. We provide a historical survey on the development of random access protocols, which culminates with the introduction of the information theoretic treatment of MRA provided by the UMAC framework.
B. We discuss recent progress in the field of MRA protocols, with emphasis on schemes that embrace the UMAC perspective. We provide a classification of some of the most

G. Liva is with the Institute of Communications and Navigation, German Aerospace Center (DLR), Wessling, Germany (email: gianluigi.liva@dlr.de).

Y. Polyanskiy is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139 USA (email: yp@mit.edu).

G. L. acknowledges the financial support by the Federal Ministry of Education and Research of Germany in the framework of the "Souverän. Digital. Vernetzt." Joint project 6G-RIC, project identification number: 16KISK022. The work of Y. P. was supported by the National Science Foundation under Grant No CCF-2131115.

promising UMAC architectures of recent introduction, highlighting their distinctive features.

C. Finally, we will outline how these theoretical developments may influence the design of future (3GPP) wireless cellular systems. In particular, we show how simple modifications [3] of the two-step random access protocol — a grant-free random access protocol that has recently been introduced in the 5G New Radio (5GNR) standard [4] — can dramatically improve its efficiency, paving the way for the support of MRA in future versions of the standard.[1]

B. Outline

The contribution is structured as follows. Section II reviews the historical progress in the theory and practice of random access protocols, discussing the blurring of the separation between the medium access control layer and the physical layer. Section III discusses the UMAC framework, highlighting the implications of merging the medium access control layer and the physical layer from a coding theory viewpoint. Emerging UMAC coding architectures are presented in IV. The grant-free random access capabilities included in the 5GNR standard are illustrated in Section V. Their limitations are discussed and possible directions for future developments are identified. Conclusions follow in Section VI.

II. From Aloha to Codes for Unsourced Multiple Access

We start with a brief perspective on the development of random access protocols. Our emphasis is on schemes developed in the context of wireless (including satellite) communications with an aim towards large-scale mMTC/IoT systems. Consequently, we will mostly ignore protocols that rely on carrier sensing, and protocols making intense use of feedback channels (such as splitting / contention tree algorithms).

This section somewhat artificially divides the evolution of random access into three periods (see also Figure 1). The first period (1970-2007) is dominated by Aloha-like schemes. The second period (2007-2017) builds on the introduction of multiuser detection (MUD) techniques to improve the performance of random access protocols. The third period (starting in 2017, and still in progress) sees a paradigm change with the introduction of the UMAC model, in which random access is viewed from a coding-theoretic perspective.[2]

A. First Period (1970–2007): Aloha and Collision Models

Initially, random access was understood as a layer-2 task, more specifically as part of the medium access control sublayer of the data link layer [8]. The introduction of Aloha [1] and of its slotted version [9] sets the ground for the development of sophisticated variations on the theme. These include carrier sense multiple access (CSMA) [10], splitting/contention-tree algorithms [11], [12] (see also [Chapter 4.3] [8]), and conflict avoiding codes [13]–[15]. Here, mutual interference among users is treated as destructive (hence, the notion of *collisions*), and protocols aim at avoiding collisions (as for CSMA), at resolving collision events via retransmissions (as in splitting/contention-tree algorithms), or controlling the number of collisions (as for conflict avoiding codes). The possibility of decoding in the presence of interference— yielding the so-called multipacket reception (MPR) capability [16]—is mostly considered in terms of *capture effect*, i.e., in contexts where user transmissions arrive at the receiver antenna with a large difference in power [9], [16], [17]. In this case the MPR capability does not stem from a receiver design based on MUD techniques [18], [19], but rather from propagation conditions in the multiple access channel. A significant departure from the collision model is represented by random access protocols that rely on spread spectrum techniques [20]–[23], where MPR is explicitly targeted by signal design. An important example of a random access protocol based on spread spectrum waveforms is the spread Aloha protocol [23], which will emerge as a high

TABLE I
List of Acronyms

5GNR	5G New Radio
AMP	approximate message passing
AWGN	additive white Gaussian noise
BAC	binary adder channel
BP	belief propagation
BS	base station
CDMA	code division multiple access
CRDSA	contention resolution diversity slotted Aloha
CS	compressed sensing
CCS	coded compressed sensing
CSA	coded slotted Aloha
CSMA	carrier sense multiple access
E-SSA	enhanced spread spectrum Aloha
FDMA	frequency division multiple access
IDMA	interleave division multiple access
IRSA	irregular repetition slotted Aloha
IoT	Internet of Things
LDPC	low-density parity-check
LTE	Long Term Evolution
MAC	multiple access
MIMO	multiple-input multiple-output
MMSE	minimum mean squared error
MRA	massive random access
mMTC	massive machine-type communication
MPR	multipacket reception
MUD	multiuser detection
OMP	orthogonal matching pursuit
PRACH	physical random access channel
PUPE	per-user probability of error
PUSCH	physical uplink shared channel
QPSK	quadrature phase shift keying
RCS	Return Channel via Satellite
SB-IDMA	sparse block interleave division multiple access
SCL	successive cancellation list
SNR	signal-to-noise ratio
SPARC	sparse regression code
SIC	successive interference cancellation
TDMA	time division multiple access
TIN	treat interference as noise
UMAC	unsourced multiple access
UT	user terminal

[1]We mention that although the paper is written with an eye towards cellular networks (3GPP), almost everything we discuss carries over without change to the *low-power wide-area networks (LP-WANs)*, such as LoRaWAN [5], mioty [6] and Zigbee [7].

[2]The choice of the dates delimiting the three periods is, of course, subjective. As a criterion, we decided to adopt the publication date of landmark papers that signal a change of perspective on the random access problem.

performance random access scheme when coupled with MUD during the second period (see the following subsection). Although the benefits of this class of protocols were immediately recognized, Aloha and CSMA dominated as the main random access techniques in wireless and wired data networks.

B. Second Period (2007–2017): Aloha-like protocols with Multi-User Detection

The second period coincides with the adoption of MUD to improve the performance of Aloha-based algorithms. In this context, a close interaction is established between the medium access control layer and the physical layer by modifying the Aloha protocol to facilitate the application of multiuser signal processing techniques. A first example is the contention resolution diversity slotted Aloha (CRDSA) protocol [24], introduced to provide efficient use of satellite return channels in machine-type applications, and currently in use by the Digital Video Broadcasting (DVB) Return Channel via Satellite (RCS) standard [25]. CRDSA relies on packet repetition and on successive interference cancellation (SIC) to improve the performance of slotted Aloha, providing tangible gains, especially at low packet error rates. A similar principle was devised in [26] in the context of contention tree algorithms. It was soon recognized that the performance of CRDSA under SIC can be analyzed by establishing an analogy with erasure decoding of low-density parity-check (LDPC) codes [27], providing the means to optimize the protocol behavior [28]. In [29] a general protocol based on the CRDSA principle — named coded slotted Aloha (CSA) — was introduced. The CRDSA scheme of [24] and the irregular repetition slotted Aloha (IRSA) scheme of [28] can be recognized as special instances of CSA. In [30] it was shown that a suitable design of the IRSA protocol allows achieving a peak throughput of one packet per slot, in the limit of large multiple access (MAC) frame lengths. Variations of CSA include the adoption of a feedback-based frameless approach [31], [32] and spatial coupling [33], as well as the elimination of the assumption of a slotted frame structure [34], [35].

In parallel to CSA techniques, MUD applied to spread Aloha represented a key development of this period. Similarly to the case of CSA, spread Aloha was studied mainly in the context of satellite mMTC/IoT networks. The enhanced spread spectrum Aloha (E-SSA) protocol [36] improves the MPR capability of spread Aloha by canceling the interference contribution of decoded packets. Due to its completely asynchronous operation, its outstanding performance and lean transmitter/receiver design, E-SSA emerged as a high-efficiency random access solution for mMTC/IoT and interactive satellite networks [37], [38].

C. Third Period (2017–): Random Access as a Coding Problem

The development of random access protocols during the first two periods has been largely based on a medium access control layer perspective. Consequently, development focused on packet-level metrics such as throughput, goodput and latency, whereas in the mMTC/IoT domain it was important to also address energy efficiency, thus requiring adequate modeling of the physical layer part. In addition, there was no model capable of capturing the fundamental aspects that differentiate random access from coordinated MAC.

The UMAC model [2] resolves both of these issues and offers an information-theoretic ground to study random access schemes. This is achieved by recasting the problem into a coding-theoretic language. We will formally introduce UMAC model in Section III, but it is instructive to first consider the following example illustrating how random access can be seen as a "coding" problem in this framework.

Example 1 (Slotted Aloha as UMAC code)**.** *Let us consider transmission with framed slotted Aloha. The frame, consisting of n complex channel uses, is divided into L slots of n_c complex channel uses each. According to the slotted Aloha protocol, an active user encodes its k-bits message into a word $\boldsymbol{w}$ of n_c symbols via an (n_c, k) block code $\mathcal{C}$, then it selects a slot to transmit the word $\boldsymbol{w}$. From a coding point of view, we can describe the encoding performed by the user as the selection of a codeword with the form*

$$\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L) \tag{1}$$

where $\boldsymbol{x}_\ell = \boldsymbol{w}$ if the user selected the ℓth slot, and $\boldsymbol{x}_\ell = \boldsymbol{0}$ (length-n_c zero vector) otherwise. Hence, the codebook realized by the slotted Aloha protocol is given by all n-tuples in the form (1) *where one block is a word from $\mathcal{C}$ and all the other blocks are zero vectors, i.e., slotted Aloha can be seen as the UMAC code*

$$\mathcal{C}_{\mathsf{SA}} = \{(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) \in \mathbb{C}^n | \boldsymbol{x}_\ell \in \mathcal{C}, \boldsymbol{x}_j = \boldsymbol{0}, j \neq \ell, \ell \in [L]\}\,.$$

Each active user transmits a codeword from $\mathcal{C}_{\mathsf{SA}}$. The cardinality of the slotted Aloha codebook is therefore $|\mathcal{C}_{\mathsf{SA}}| = L|\mathcal{C}|$. The selected slot can either be random (as in the original slotted Aloha) or it could be chosen by computing a hash function of the payload data, thus providing extra parity checks for decoder.

As the example shows, in the UMAC framework, the "code" incorporates aspects of both the physical layer and the medium access control layer. We found that presenting this example is crucial for explaining the idea of UMAC codes to network engineers, since they traditionally considered codes to operate only at the physical layer, and not think of the medium access control protocol part as "code". However, as we can see in the case of slotted Aloha, the selection of the slot used for the transmission can be thought of as rudimentary code. The inclusion of control layer aspects in UMAC is an essential step in building a unifying theory of MRA protocols, allowing a fair comparison of several MAC strategies.

With this preview, we are ready to introduce the UMAC model (Section III) and discuss emerging practical approaches (Section IV).

III. Random access from an information-theoretic perspective

Let us start with describing the standard MAC setting as studied in communication theory and information theory:

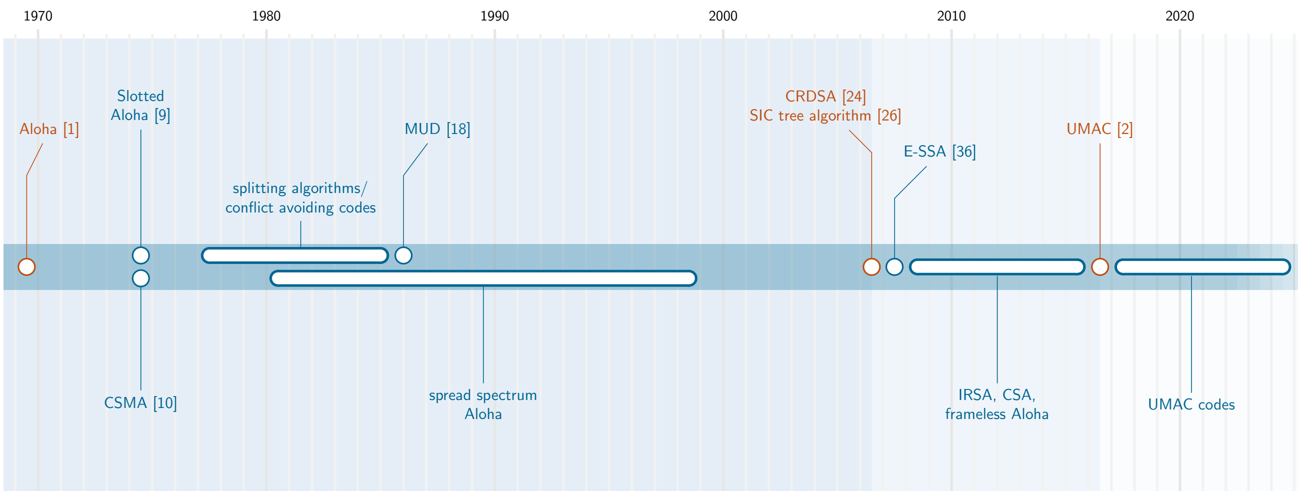


Fig. 1. Timeline of the development of random access protocols.

- *Uplink:* A single BS transmitting beacons and listening on a common broadcast channel for the uplink transmissions.
- *Users:* Multiple users K (K at most a few hundred) are simultaneously using the uplink channel. The identities of the communicating users are known to the BS (due to prior control plane exchanges).
- *Payload:* The users are sending either continuous streams (voice communication) or a large number of information bits in each session (data transfer).
- *Multiple access:* Before the users are allowed to join the uplink communication, they have to announce their existence to the BS via a different channel — known as physical random access channel (PRACH) in the Long Term Evolution (LTE) standard. Upon establishing their communication intent, the BS instructs all currently active users on how they should share the channel access (in LTE the BS schedules resource blocks and configures time-offsets). The resulting allocations are distributed to the users as part of the beacon broadcast.

We note that in this setting the uplink multiple access is completely *coordinated* by the BS. Because each user is sending a very large payload, the overhead that it spent on the resource acquisition phase and coordination are amortized. From an information-theoretic point of view, thus, the main challenge in the setting above is that of designing K (different!) channel codes, such that when K random codewords (one codeword from each) are transmitted simultaneously on the uplink, the BS is able to recover each of them with a high probability of success. How to choose these K codes is the subject of the information-theoretic MAC, see [39, Section 15.3]. However, for the additive Gaussian noise channels, it is known that using a single standard (point-to-point) error-correcting code and simply allocating non-overlapping time-frequency resources to different users is optimal from the channel capacity point of view.

Next, let us describe the mMTC/IoT communication setting. Specifically, we will consider the following assumptions:

- *Uplink:* A single base station (BS) transmitting beacons and listening on a common broadcast channel for the uplink transmissions.
- *Users:* A very large (order a million) number of users (mMTC/IoT devices), of which majority are idle. When idle, users do not monitor the BS transmissions to conserve battery. Despite a large total number of users, only K_a of them have any data to send on the uplink (K_a again is on the order of a few hundred).
- *Payload:* When users have data to send, their messages are rather short (100s of bits). The message payload may simply be an identity (cryptographically signed) of the user, or the identity plus a short status update.
- *Random access:* Users wake up from the idle state at random, without BS knowing who is awake at any given time. Since the duty cycle (the amount of time the user has to stay on before completing its radio access) directly contributes to battery depletion, the desire of each users is to initiate communication immediately after waking up.

Comparing the two settings side by side clearly shows the salient feature of the latter: While the total number of communicating users may be roughly the same ($K \approx K_a$), the identities of communicating users in the second setting are unknown. Correspondingly, the communication needs to proceed in a completely uncoordinated way. How is it possible to achieve any reliable data transmission by multiple users without them coordinating in some fashion?

The simplest and most ubiquitous solution is the Aloha protocol: Each user, whenever it has data to send, simply transmits its message on the uplink (the CSMA variation requires the user to first check that the uplink is idle and, if not, to retry at a random later time). We see that if transmissions of the users are very short and wake-up times are random, the chance of collisions is low. Let us try to estimate this chance.

Suppose that at the beginning of a frame (following a beacon) we have K_a users ready to transmit their data. We can slice our frame into L nonoverlapping slots[3]. In accordance

[3]Instead of this TDMA-like idea, we could also divide the channel according to some other orthogonal basis. For example, for random access in LTE (PRACH) the users choose from $L = 64$ possible Zadoff-Chu sequences, which all overlap in time but otherwise are orthogonal.

with the Aloha principle, each of K_a users selects one of the L slots at random and places its message there. In this case, let us fix one of the users and ask what is the probability that someone else selects its slot for communication:

$$P_{\text{collision}} = 1 - \left(1 - \frac{1}{L}\right)^{K_a - 1} \le \frac{K_a - 1}{L}\,.$$

This calculation implies the following important conclusion: Unless we are able to decode the packets that collide (are transmitted in the same slot) there is an error floor $\approx \frac{K_a-1}{L}$ for the probability of recovering a user's message. Since making L very large is impractical, we are led to the natural conclusion that any random access scheme must use some MPR capabilities.

Let us summarize our findings. For realistic values of L, the Aloha principle alone is not capable of producing an uncoordinated random access with low probability of error. All of the users who selected the same slot (or the same Zadoff-Chu preamble in LTE's PRACH) appear completely identical to the BS. That is, from the point of view of the BS, these users are all employing an identical transmission strategy, with the only difference that each of them is transmitting a different payload. How to produce such a transmission strategy is precisely the topic of *UMAC coding theory* (with "U" standing for uncoordinated or unsourced).

We now describe more formally what a UMAC code is meant by. First, we define the two channel models that we use in the remainder of the paper. Second, we recall the definition of a point-to-point error-correcting code. Third, we define the UMAC code.

Channel models. We only consider single antenna channels. Recall that a K-user additive white Gaussian noise (AWGN) channel with K users and blocklength n takes as input K vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K \in \mathbb{C}^n$ and outputs a random $\boldsymbol{Y} \in \mathbb{C}^n$ according to

$$\boldsymbol{Y} = \sum_{i=1}^{K} \boldsymbol{x}_i + \boldsymbol{Z}, \qquad \boldsymbol{Z} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$$

where $\boldsymbol{Z}$ is a complex vector with n i.i.d. complex normal entries of power σ^2 per entry. All channel inputs are subject to a power constraint $P > 0$, that is we must have

$$\|\boldsymbol{x}_i\|^2 \le nP \qquad \forall i \in \{1, \ldots, K\}\,.$$

The quasi-static Rayleigh fading channel is defined similarly, except that each user's input is scaled by an independent channel gain H_i , that is we have $\boldsymbol{Y} \in \mathbb{C}^n$ generated as

$$\boldsymbol{Y} = \sum_{i=1}^{K} H_i \boldsymbol{x}_i + \boldsymbol{Z}, \qquad H_i \overset{i.i.d.}{\sim} \mathcal{CN}(0, 1)\,.$$

See [40, Sections 20.3 and 20.9] for more on these channel models. Note that since we are primarily interested in the uncoordinated case, we cannot assume that H_i's are known at the receiver.

Point-to-point channel codes. A special case of $K = 1$ is called a point-to-point channel (since there is only one transmitter and one receiver). An (n, M, ϵ) error-correcting code for a point-to-point channel is defined as a collection of M codewords $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M$ together with a decoder function $g : \mathbb{C}^n \to [M] := \{1, \ldots, M\}$ such that the average probability of error satisfies

$$P_e = \frac{1}{M} \sum_{m=1}^{M} \mathbb{P}[g(\boldsymbol{Y}) \neq m | \boldsymbol{X} = \boldsymbol{c}_m] \le \epsilon\,.$$

See [40, Chapter 17] for more on the definition of the error-correcting codes.

It is known that for the AWGN channel the best point-to-point codes satisfy [41]:

$$\begin{aligned} \log M &= nC - \sqrt{nV} Q^{-1}(\epsilon) + O(\log n) \\ &\approx nC - \sqrt{nV} Q^{-1}(\epsilon) \end{aligned}$$

where $Q^{-1}(\cdot)$ is the inverse of the Q-function, $C = \log(1 + \frac{P}{\sigma^2})$ is the channel capacity and $V = \frac{P(P+2\sigma^2)}{(P+\sigma^2)^2} \log^2 e$ is the channel dispersion.

Besides probability of error, another important figure is the normalized energy-per-bit, or $\frac{E_b}{N_0}$, defined as

$$\frac{E_b}{N_0} := \frac{nP}{2\sigma^2 \log_2 M}$$

which quantifies the energy spent by a user terminal to transmit an information bit. See [40, Section 21.1] for more on energy-per-bit.

UMAC codes. Finally, we define the new type of error-correcting codes that will enable a principled exploration of random access with MPR. An (n, M, ϵ, K_a) UMAC code is a collection of M codewords $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M \in \mathbb{C}^n$ and a decoder function $g : \mathbb{C}^n \to \binom{[M]}{K_a}$, where $\binom{[M]}{K_a}$ denotes a collection of all subsets of K_a elements from the set $[M]$. The codebook and the decoder should satisfy the *per-user probability of error (PUPE)* $\le \epsilon$ constraint. To define PUPE we suppose that user j selects uniformly at random (independently of other users) a message W_j from $[M]$ and sets its channel input $\boldsymbol{x}_j = \boldsymbol{c}_{W_j}$. Once all K_a channel codewords are selected they are input to a K_a-user channel (AWGN or fading), which produces the output $\boldsymbol{Y}$. The decoder output $g(\boldsymbol{Y})$ is the list of messages that the decoder believes were transmitted by the users. The PUPE is defined as[4]

$$\mathsf{PUPE} := \frac{1}{K_a} \sum_{j=1}^{K_a} \mathbb{P}[W_j \notin g(\boldsymbol{Y})]\,.$$

That is, the PUPE measures the probability that a user's message is going to be absent from the list of messages decoded by the receiver. See [2] for more formal details on UMAC codes, as well as connections to related concepts in combinatorics and sparse regression.

Let us reflect on some of the ideas encoded in the mathematical definition above. First and foremost, despite having K_a active users, there is only one codebook shared by all of them. This requirement formalizes the situation in which uplink transmission happens in an uncoordinated way (the only

[4]More exactly, to match the definition in [2], we have to include in the error event also the case that W_j clashes with a message W_i for some $i \neq j$. However, since the chance of this happening is at most $\frac{K_a^2}{2M}$ and we are focused on $M = 2^{100}$ in this survey, we prefer to omit this irrelevant term.

coordination is a common frame boundary signalled by the BS' beacon).

Second, while traditional error-correcting codes are required to arrange their codewords in a way that allows the identity of a transmitted point to be decodable from a noisy observation, the UMAC code faces a more difficult challenge: a subset of any K_a codewords should be decodable from observing a noisy sum of those codewords. In particular, since the users all employ the same codebook and the channel is invariant to permutation of the users, we can see that the decoder is not able to ever associate messages to users and is only recovering an unordered *subset* of codewords. This is the reason for the name *unsourced* MAC, since the messages are not sourced back to their originators. We remark, however, that as identity of the user is likely a part of the payload, this ambiguity is easy to resolve at higher layers. Leaving the problem unsourced, though, makes the coding-theoretic part cleaner and more natural.

Third, we observe an important departure from the classical MAC in information theory: the PUPE criterion only bounds the probability of error for an individual user. This choice is not only reasonable from the system-level point of view (since the uplink design criterion is to satisfy a certain accuracy for each separate user), but also natural theoretically. If an error is declared whenever any of the K_a messages is misdecoded (as is done classically), the required E_b/N_0 grows without bound as K_a increases, cf. [42, Slide 83].

In summary, the UMAC code is defined in a way that formalizes the notion of uncoordinated random access by a-priori indistinguishable users and defines error probability in a way that allows analysis even for large K_a without penalizing energy efficiency.

IV. Prominent Unsourced Multiple Access Architectures

Quickly after the introduction of the UMAC setting, several coding schemes were proposed to approach the UMAC performance limits, over the Gaussian MAC as well as over the fading MAC. Most of the constructions are directly inspired by the compressed sensing (CS) perspective adopted in [2], and can be classified according to four emerging architectures, namely: Aloha-based schemes enhanced by some form of MPR capability, coded compressed sensing (CCS) schemes, preamble-based architectures, and spreading-based architectures. The different schemes are discussed in the following subsections. Given the rapidly evolving landscape of UMAC code constructions, no attempt will be made to provide a thorough review of the existing methods. Rather, the focus will be on the distinctive features of the different architectures. For some selected schemes, results on the Gaussian MAC with $n = 30000$ channel uses are reported in Figure 5.

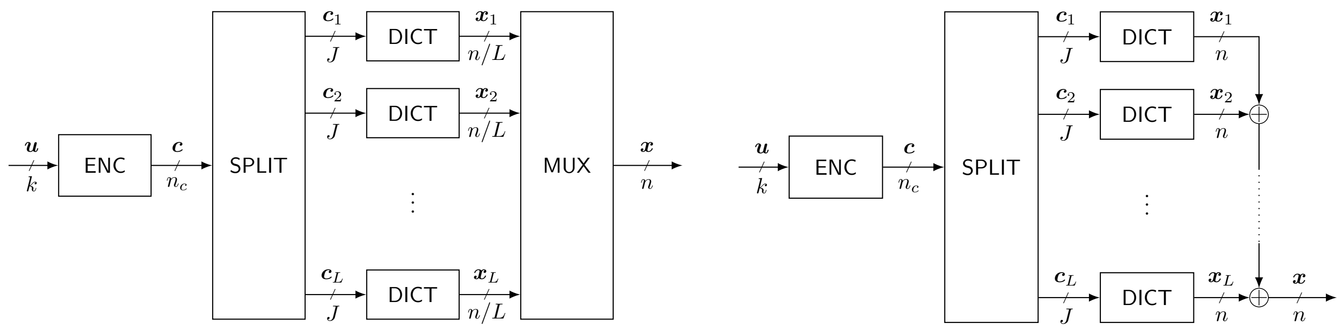
A. Multi-Packet Reception slotted Aloha Architectures

Although the concept of MPR in Aloha systems is relatively old (see Section II-A), the design of coding mechanisms that allow the decoding of moderate-size collision sets has received renewed interest in the UMAC context. The use of slotted Aloha as a basis for developing powerful UMAC schemes stems from the following observation: Decoding collisions that involve many users according to the model of Section III entails high complexity. On the contrary, moderate-size collision clusters can be resolved with low complexity using powerful error-correcting codes. In slotted Aloha, the average collision set size in a slot is a fraction of the number of active users—equal to the number of active users, divided by the number of slots in which the MAC frame is partitioned. This enables the use of effective strategies to resolve multiple collisions in a slot.

The design of UMAC schemes relying on this principle was introduced in [43], where a T-fold Aloha construction was introduced. The construction implements a layered approach, where each message is encoded first by an outer code for T-users binary adder channel (BAC), and then by an inner binary linear code, designed to enable the decoding of the modulo-2 sum of the colliding codewords. The outer code can be based on the columns of the parity-check matrix of a T-error-correcting BCH code, while the inner code can be any powerful short-blocklength linear block code with low decoding complexity. The inner code decoder delivers with high probability the module-2 sum of the colliding codewords to the outer code decoder. If the cardinality of the collision set is within the decoding radius T of the outer code, the set of collision messages is resolved. Otherwise, the collision batch is lost.

Effective MPR mechanisms in T-fold Aloha schemes for the Gaussian MAC channel have been proposed in [44], [45], which rely on joint user decoding using LDPC and polar codes, respectively. Both constructions also include packet repetition, allowing interference cancellation across slots in a way that is reminiscent of the IRSA protocol [28]. The performance of the scheme of [45] is shown in Figure 5. A T-fold Aloha construction was introduced in [46], where the MPR capability is provided by using short, terminated convolutional codes with joint decoding over a super-trellis tailored to the collision cluster. After convolutional encoding, randomized signature sequences are applied to the encoded packet. Their role is to enhance the performance of the multi-user trellis decoder.

T-fold Aloha schemes also yield very promising results in fading channels. Joint user decoding and channel estimation on the quasi-static Rayleigh fading channel have been investigated in [47]. The approach, which is based on LDPC codes with multi-user belief propagation (BP) decoding, allows supporting moderate-size user populations with single-antenna receivers. A largely improved performance was achieved in [48] by replacing LDPC codes with low rate polar codes, with a simple treat interference as noise (TIN)-SIC decoder architecture. In fact, assuming independent fading coefficients, user separation turns out to be an easier task, allowing the use of low-complexity strategies based on single-user decoding and interference cancellation. A distinctive feature of the approaches of [47], [48] is the lack of pilot sequences, with channel estimation performed directly on the transmitted data. A pilot-based T-fold Aloha scheme for massive multi-antenna systems was proposed in [49], where the adoption



(a) CCS: $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_L)$.

(b) CCS, in the SPARC variant: $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \dots \boldsymbol{x}_L$.

Fig. 2. Coded compressive sensing architectures.

of randomized pilot hopping patterns, interleaved with the segments of the user codewords, was used to mitigate the pilot contamination effect that stems from the impossibility to allocate orthogonal pilot sequences to all users. In [50] a related approach was introduced, employing sparse transmission patterns such as in the frameless Aloha [31], [32] protocols. In [51], [52], a slotted Aloha scheme based on the construction of [53] was proposed. Here, the pilot sequences are selected by the users from a common sensing matrix, and CS techniques are used to obtain estimates of the channel coefficient vectors of the detected users. This class of schemes was shown to support a large number of active users with a signal-to-noise ratio (SNR) that is only marginally larger than the one required in the single-user case.

B. Coded Compressive Sensing Architectures

Recognizing the elegance of the perspective outlined in [2], efforts to construct practical UMAC schemes based on the CS analogy were initiated in [54], [55] by introducing CCS architectures. CCS uses a divide-and-conquer approach to address the exceedingly large dimensions of the sparse recovery problem associated with UMAC. The principle underlying CCS is to divide the message in L small blocks of J bits each, and to encode each block through a CS dictionary of cardinality 2^J, producing encoded blocks of length n/L. With J in the order of a few bits, it is possible, at the receiver end, to apply low-complexity sparse recovery algorithms working on the $(n/L) \times 2^J$ sensing matrix obtained by stacking the 2^J dictionary sequences. Assuming a perfect recovery of the transferred blocks, the decoder is left with the task of "stitching" together the blocks associated with the individual user transmissions. The task can be accomplished by treating the outputs of the sparse recovery phase as observations of a so-called A-channel [56] (see [57] for an elegant casting of CCS as a concatenated transmission scheme). Therefore, the code constructions for the A-channel can be used to reconstruct the individual user messages. For this purpose, in [54] an ingenious construction of tree codes was proposed. A simplified description of the CCS architecture as defined in [54] is shown in Figure 2a.

CCS in combination with sparse regression codes (SPARCs) [58], [59] was proposed in [57]. Similarly to the approach of [54], each message is divided into L blocks of J bits each. Unlike [54], the L blocks are encoded using a superposition code. According to the SPARC construction, an $n \times L2^J$ matrix is partitioned into L submatrices with dimensions $n \times 2^J$. The columns of each submatrix define a local dictionary that is used to encode a J-bits block. The encoded blocks are then added together. In [57], a modified approximate message passing (AMP) decoder was introduced to extract the message blocks transmitted by the users, with the tree code of [54] used to reconstruct the user messages. The corresponding CCS architecture is illustrated in Figure 2b. The performance on the Gaussian MAC of the original CCS scheme of [54] (enhanced by a successive cancellation procedure, whereby decoded packets are subtracted from the signal at the input of the CS recovery algorithm) and the one of the SPARC-based construction of [57] are shown in Figure 5. Remarkably, the latter shows a minimal increase of the required SNR was the number of users grows, up to 150 active users. It was soon recognized that the architecture of Figure 2b is more general and it comprises the original architecture depicted in Figure 2a as a special case, where the sensing matrix is organized in a block diagonal form [60].

A SPARC-based CCS architecture with a modified outer code to allow joint decoding of the inner CS code and of the outer code via iterative AMP/BP was proposed in [61]. The approach allows gains in the order of a few tenths of a dB over the scheme of [57] in the moderate channel load regime. Nonbinary LDPC codes as outer codes with iterative BP/AMP decoding and denoising were proposed in [62]. The use of an outer binary LDPC code exploiting soft outputs delivered by a modified AMP decoder has recently been explored in [63] in a massive multiple-input multiple-output (MIMO) setting, supporting a very large number of active users.

C. Preamble-based Architectures

Similarly to CCS schemes, preamble-based architectures implement an attack to the UMAC problem by a divide-and-conquer approach. The idea is to apply a CS-based detection of active users by means of preambles that users select from a moderate-size dictionary, thus, enabling low-complexity detection. The preambles are used to *announce* the resources that will be used for data transmission in a subsequent phase. The principle is described in generic terms in Figure 3. A first example of this class of UMAC schemes was introduced in [64],

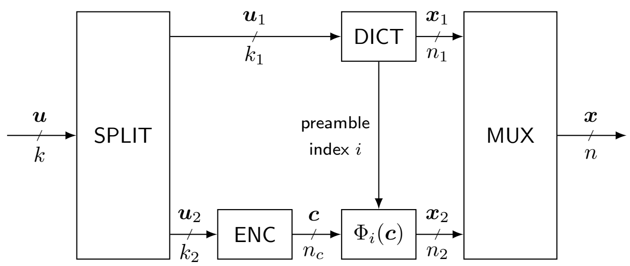


Fig. 3. Preamble-based: $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$.

where the preambles were associated with interleave division multiple access (IDMA) access patterns [65]. The scheme, named sparse IDMA, works as follows. User messages are divided into two parts. A first part, composed by k_1 bits, is used to address a dictionary of 2^{k_1} preambles. Denoting by $\boldsymbol{x}_1$ the selected preamble and by i the corresponding index in the preamble dictionary, the index is used to select a repetition-and-interleaving pattern Φ_i (see Figure 3). The remaining k_2 bits of the message are encoded by means of a (n_c, k_2) binary linear block code, resulting in the codeword $\boldsymbol{c}$. The selected repetition-and-interleaving pattern Φ_i is applied to $\boldsymbol{c}$ generating a sparse vector $\boldsymbol{x}_2$ that is appended to the preamble $\boldsymbol{x}_1$. At the receiver side, the detection of transmitted preambles is performed through a suitable CS algorithm. Given the list of detected preambles, an IDMA decoding algorithm is employed to resolve the transmissions associated with the second part of the access frame.

The principle can be generalized to the use of strategies different from IDMA for the second phase of transmission. In this sense, the mapping Φ_i in Figure 3 can be interpreted as a general transformation of the word $\boldsymbol{c}$. In [66], the mapping involves repetition, interleaving, and multiplication of the resulting vector by a binary signature, with the introduction of a randomized transmission delay. A similar construction, optimized for the block-fading Rayleigh channel, was introduced in [67]. Both in [64] and in [67], sophisticated multiuser decoding strategies are used to decode the second part of the detected transmissions. In [64], joint decoding is performed over the factor graph that couples the transmissions, according to the IDMA patterns identified by the preambles. In [67], iterative MUD detection with soft interference cancellation is employed before decoding. The performance of the sparse IDMA scheme of [64], where the (n_c, k_2) binary linear block code is a binary LDPC code optimized for the multiuser setting, is depicted in Figure 5. The scheme works withing 2 dB from the achievability bound up to 100 active users.

A scheme closely related to sparse IDMA was recently introduced in [68], where the mapping Φ_i simply spreads the n_c codeword bits over n_c locations defined by the preamble index i. The scheme of [68] can be operated over the Gaussian MAC without the transmission of the preamble, i.e., with a detection of the used access patterns that is performed through a likelihood ratio test that treats the codeword symbols as i.i.d. random variables in $\{-1, +1\}$. Nevertheless, to facilitate the detection of the access sequence and to enable the estimation of the channel on fading channels, the transmission of the selected preamble is considered in [69]. Remarkably, the scheme of [69] closely approaches the achievability bound of [2] (see Section III) on the Gaussian MAC channel in the moderate load regime with a lean TIN-SIC receiver architecture, also thanks to the use of polar codes (concatenated with an outer CRC code).

We will see in Section V that the grant-free mechanism recently introduced in the 5GNR standard adopts elements of the preamble-based architecture.

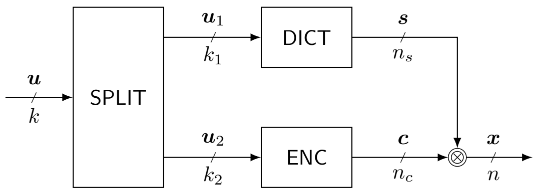
D. Spreading-based Architectures

The excellent match between random access and code division multiple access (CDMA) and, more generally, spread spectrum techniques has been widely recognized, and it has been the basis of some of the most advanced random access protocols developed prior to the introduction of the UMAC framework: the spread Aloha protocol [23]. Spread Aloha employs a unique spreading sequence for all users and relies on time asynchronism among users to facilitate detection and decoding. By tailoring the protocol to the UMAC setting, it has been shown recently that spread Aloha with TIN-SIC decoding can closely approach the achievability bound on the Gaussian MAC channel [70].

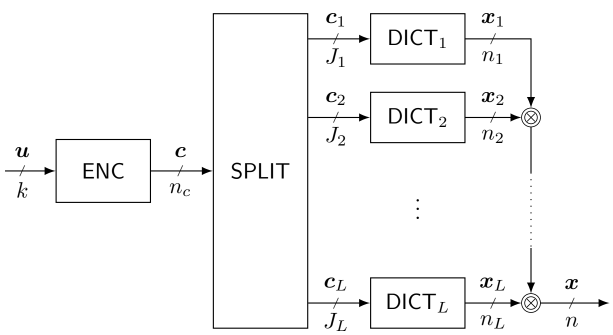
Elegant spreading-based architectures explicitly designed for the UMAC were introduced in [71], [72]. Regarding Aloha spread, both solutions are designed assuming perfect synchronization to the UMAC framing structure, and rely on the use of multiple data-dependent spreading sequences.

The construction of [71] is outlined in Figure 4a. Following the approach of preamble-based architectures, the user message is divided into two parts. A first part ($\boldsymbol{u}_1$, composed by k_1 bits, see Figure 4a) is used to select a spreading sequence from a dictionary. The second part of the user message is ($\boldsymbol{u}_2$, composed by k_2 bits) is encoded with a (n_c, k_2) binary linear block code. Direct sequence spreading with the selected spreading sequence is applied to the codeword symbols. At the receiver side, energy detection is used to extract the set of spreading sequences that were selected by the active users. Decoding proceeds by linear minimum mean squared error (MMSE) data estimation, followed by decoding and SIC. The performance of the scheme on the Gaussian MAC, with a polar code used to encode the second part of the message, is shown in Figure 5. A remarkably small gap (less than 0.5 dB) from the achievability bound can be observed, up to 100 active users. Constructions similar to the one of [71] were introduced in [73], where sparse spreading codes were used, and in [74], where the original scheme [71] was engineered to operate on quasi-static fading channels with massive MIMO arrays.

The spreading-based architecture introduced in [72] follows a less conventional path. In particular, a recursive encoding mechanism is used, with L stages of spreading. The approach works as described in Figure 4b: the information message is first encoded by an outer code, resulting in the outer codeword $\boldsymbol{c}$, which is then divided in L blocks $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_L$, where the ith block comprises J_i bits. Each block is encoded through a local dictionary (referred to as *sub-constellation*), which is



(a) Spreading-based: $\boldsymbol{x} = \boldsymbol{c} \otimes \boldsymbol{s}$.



(b) Spreading-based (tensor product): $\boldsymbol{x} = \boldsymbol{x}_L \otimes \boldsymbol{x}_{L-1} \otimes \cdots \otimes \boldsymbol{x}_1$.

Fig. 4. Spreading-based architectures.

possibly different for each of the L blocks. Denoting by $\boldsymbol{x}_i$ the local dictionary sequence that encodes $\boldsymbol{c}_i$, the output of the encoder is the tensor product of the sequences $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_L$. Joint decoding and channel estimation are performed on the decoder side, using rank-1 tensor decomposition to perform user separation. Single-user decoding follows for the detected users. The beauty of the architecture devised in [72] comes from its ability to separate users without relying on additional pilot sequences. The construction provides a competitive performance over block fading channels, with and without multi-antenna receivers.

E. Maturity

The four architectures described in the previous subsections depart from classical random access protocols that are found in many existing wireless communication systems. In this section we wanted to loosely rank the new architectures in terms of how ready they are for real-world deployment, in our opinion.

The vast majority of new architectures makes some use of CS to solve part of the decoding problem: a few MPR Aloha schemes employ CS to detect pilot sequences of colliding users, CCS builds on CS to deliver an A channel to the outer tree code, preamble-based architectures use CS techniques to identify preambles of active users, and certain spreading-based architectures use CS to detect the spreading sequences selected by users. Most architectures foresee a layered approach to decoding. For example, in CCS the receiver decodes using a serial approach that is typical of concatenated coding schemes: an inner decoder (based on CS) outputs a set of sub-blocks associated to each slot, and an outer decoder has the task to stitch together the sub-blocks that form the message of each active user. In preamble-based architectures, the receiver first detects the preambles selected by the active users. This information is then used to direct a second decoding phase, which closely resembles the decoding of a coordinated non-orthogonal MAC transmission scheme. It is certainly meaningful to refer to architectures that limit the differences to canonical random access protocols as more "mature", from an application viewpoint. Following this principle, MPR-based Aloha architectures that rely on non-orthogonal pilot fields require very minor modifications to Aloha-based systems. Similarly, spreading-based architectures are natural candidates to improve the performance of spread Aloha systems. We will see in Section V that preamble-based architectures are closely related to the random access protocols used by the 5GNR standard. These three architectures can be considered to be relatively mature. CCS, on the contrary, relies on a more disruptive approach, which does not have any counterpart in adopted random access protocols. Therefore, the maturity of CCS can be considered relatively low.

The maturity of the different approaches is also strongly related to the decoding / detection techniques required to harvest the performance gains. Schemes that rely mostly on single-user receiver chains — possibly aided by interference cancellation — exhibit a high level of maturity. In fact, SIC-based receivers for random access protocols have been widely adopted in satellite communications [25], [37]. On the contrary, joint multiuser decoding (either in the form of joint LDPC or polar decoding, or in the form of tree code decoding) is less explored in practical implementations.

Based on these considerations, we may provide a high-level discussion of the level of maturity for some of the schemes discussed in the previous subsections. Introducing a binary classification that focuses on the upcoming 3GPP standardization efforts — "6G-ready" for schemes that present a high level of maturity, and "beyond-6G" for schemes that may require a longer engineering phase — we summarize the discussion as follows.

MPR slotted Aloha architectures. Schemes that rely on TIN-SIC decoders with slotted Aloha, using non-orthogonal pilot sequences for channel estimation [44], [49]–[53] can be considered 6G-ready. Schemes that rely on joint multiuser decoding [47], [48] require further investigations in terms of hardware architecture, and hence can be categorized as beyond-6G.

CCS architectures. CCS architectures rely on a construction that is largely unexplored, in practical implementations. For this reason, we believe that CCS represents a beyond-6G technology.

Preamble-based architectures. Preamble-based architectures can be considered in general as quite mature. A distinction shall be made between schemes that require joint multiuser decoding [64], [67], and schemes that rely on single-user decoders, aided by SIC [69]. While the latter are 6G-ready, the former should be classified as beyond-6G. We will nevertheless see in Section V that schemes closely resembling sparse IDMA [64] can be re-engineered to work with single user decoders (notably, largely reusing blocks already part

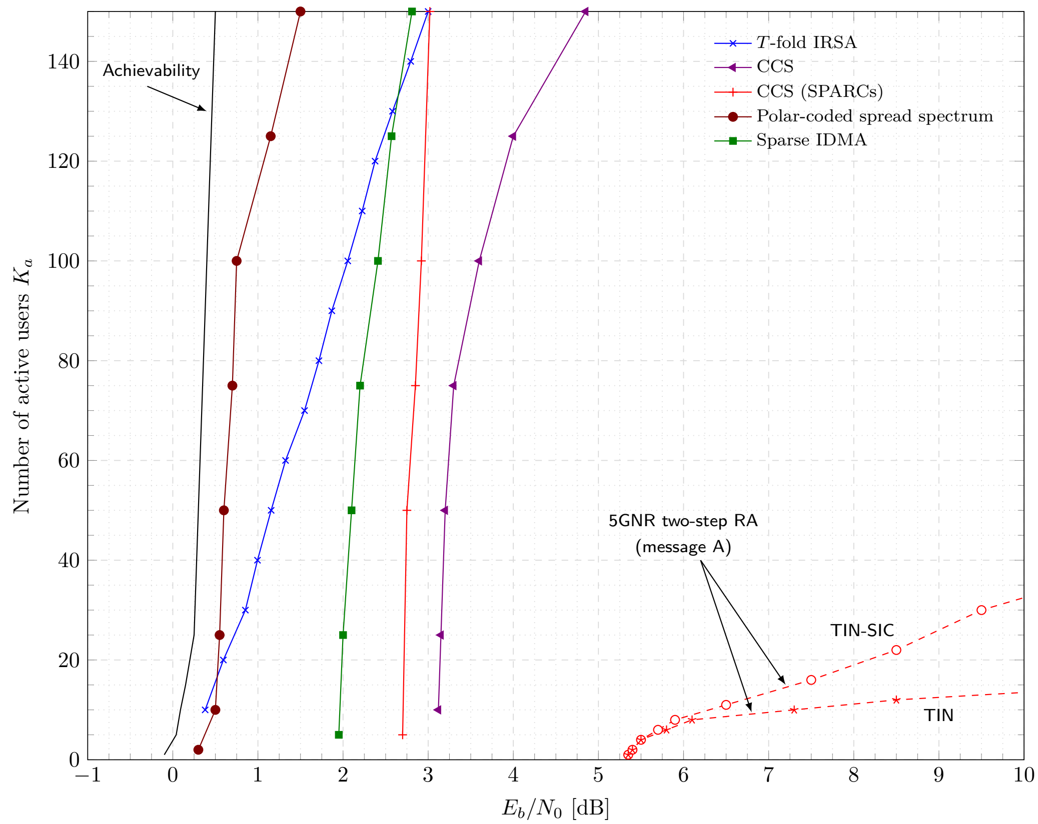


Fig. 5. Minimum SNR required to achieve a $\mathsf{PUPE} = 5 \times 10^{-2}$, over the Gaussian MAC channel. The frame length is $n = 30000$ real channel uses. For the 5GNR two-step random access protocol, $n = 32556$.

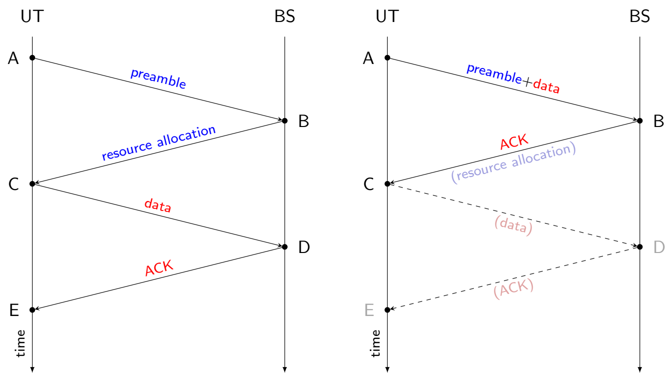
of the 5GNR standard), retaining or even improving the remarkable performance of the original system, rendering sparse IDMA a 6G-ready technology.

Spreading-based architectures. The spreading-based construction of [72] performs joint decoding and channel estimation, exploiting rank-1 tensor decomposition to perform user separation. The receiver architecture is here highly innovative and sophisticated, and it is unexplored in system implementations. We classify the scheme as beyond-6G. On the contrary, upon retrieving the set of used spreading sequences, the construction of [71] reduces to a synchronous CDMA scheme, whose implementation can be based on the numerous developments of CDMA systems. We thus consider the spreading-based architecture of [71] as 6G-ready.

V. Grant-Free Access in Cellular Networks

In the LTE (including Narrowband IoT) [75], [76] and the 5GNR 3GPP standards [77], random access is based mainly on a four-step handshake between the user terminals and the BS. The approach falls under the category of *grant-based* random access protocols. Its behavior is illustrated in Figure 6a, and it begins with the transmission of a preamble by the user terminal. In the 5GNR standard, the preamble is randomly chosen from a dictionary formed by 64 Zadoff-Chu sequences, with a sequence length that can be set to 139 or 839 complex values (the preamble length and, possibly, preamble repetition policies are defined by the network configuration). Preamble transmission takes place in random access slots (PRACH), where simultaneous preamble transmissions may collide. At the base station, preamble detection is performed, and feedback is sent to user terminals, which includes a resource allocation — in the form of a *physical uplink shared channel* (PUSCH) *occasion* — for each detected user. The user terminals can then proceed with the transmission of their data packets in the assigned resource units. The base station finally acknowledges the correct reception of the packets.

With Release 16 of the 5GNR standard, a new random access procedure has been included, which incorporates elements of *grant-free* access schemes [4]. The protocol, commonly known as *two-step random access*, was introduced with the main objective of reducing the delay entailed by the legacy four-step approach. The 5GNR two-step random access protocol works according to the preamble-based architecture described in IV-C. More specifically, a set of N resources (PUSCH occasions) is associated with the 64 preambles from the same dictionaries used by the four-step procedure. Both a one-to-one mapping (with preambles pointing distinct resource units) and many-to-one mappings (with multiple preambles pointing to the same resource unit) are possible. User terminals initiate their transmission by selecting a preamble to be transmitted within a random access slot. Each preamble "announces" the PUSCH occasion that will be used for data transmission by the user terminal, as detailed in Figure 7. The first step (referred to as *message A* transmission in the



(a) Four-step (b) Two-step

Fig. 6. Random access procedures employed by LTE/5GNR standards. (a) Four-step random access: the user terminals (UT) transmit a preamble (A). Upon detection of the transmitted preambles, the base station (BS) provides a resource allocation to each detected user (B). UTs transmit their data packets in the allocated resources (C). The BS acknowledges correctly decoded packets (D). The procedure ends when the UT receives the acknowledgement (E). (b) Two-step random access (Release 16 of the 5GNR standard): A UT transmits a preamble, that directly points to the resource that will be used to transmit the data packet. The transmission of data packets follows without waiting for a resource allocation (A). At the BS, preambles are detected and decoding is attempted in the resources pointed by the preambles. For detected UT transmissions, an acknowledgment is sent to the UTs that are successfully decoded (B). For detected UT transmissions that do not result in successful decoding, orthogonal resources are allocated for the retransmission of the data packet, resuming the four-step random access procedure.

standard) is completed by transmission of the packet in the selected resource unit. The base station performs preamble detection and attempts decoding of the packets transmitted within the PUSCH occasions that were signaled by the detected preambles. If decoding succeeds, the two-step procedure ends with the base station acknowledging the correct reception to the user terminal (*message B* transmission in the 5GNR jargon). If decoding fails, 5GNR two-step random access protocol allows to resume the four-step procedure, i.e., a negative acknowledgment if sent to the user terminal, contextually with a resource allocation.

Both the four- and the two-step random access protocols envisaged by the 5GNR standard are clearly not designed to support massive IoT networks. This observation stems from the very limited number of preambles, which is almost two orders of magnitude smaller than the cardinality of the dictionaries adopted by more advanced schemes based on the same preamble-based architecture (see Section IV-C).

With the perspective of equipping future releases of the 3GPP standard with a fully-fledged grant-free MAC protocol, and recognizing that standard updates favor solutions that have a limited impact on the system architecture, a sensible question is whether the message A transmission phase of the two-step random access protocol—or modifications of it—can be used to support massive connectivity scenarios. This critical question has recently been addressed in [3]. The next subsection reviews some of the conclusions of that study.

A. Two-Step Random Access: Performance

Figure 5 shows the performance of two-step random access (message A transmission) over the Gaussian MAC. The analysis is given for a specific configuration of the scheme: The short Zadoff-Chu preamble family is used, with 64 sequences of length 139 (complex channel uses), repeated twice (according to the A1 short preamble configuration of the 5GNR standard, see [78, Chapter 16]). The data packet encodes 100 information bits in 500 codeword bits, which are then mapped ontoquadrature phase shift keying (QPSK) symbols, resulting in 250 channel uses. The $(500, 100)$ 5GNR LDPC code (from base matrix two) is used for encoding [79]. Taking into account $N = 64$ available PUSCH occasions, the frame length is $n = 2 \times 139 + 64 \times 250 = 16278$ (equivalent to 32556 real degrees of freedom). At the receiver side, the orthogonal matching pursuit (OMP) algorithm [80], [81] is used to detect preamble transmissions. The decoding of a packet is attempted at the PUSCH occasions pointed by the detected preambles. The performance is evaluated with and without the aid of SIC, which, when present, is again assumed to be ideal. The case where no SIC is applied is referred to as TIN, whereas TIN-SIC referrers to the application of interference cancellation. When SIC is employed, preamble detection is repeated after canceling the interference contribution caused by the preambles associated with successful decoding attempts.

At low channel loads, the minimum SNR required to achieve the target $\mathsf{PUPE} = 5 \times 10^{-2}$ is largely influenced by the energy overhead introduced by the preamble (approximately 3.2 dB), and by the inherent suboptimality of the $(500, 100)$ 5GNR LDPC code, which exhibits a loss of ≈ 1.6 dB with respect to single-user finite blocklength achievability bounds. Note that the first phase of the protocol behaves as slotted Aloha, with the addition of the preamble to announce PUSCH occasions used for transmission. A classical slotted Aloha receiver would proceed by trying to decode each of N resource units — possibly, by preceding the actual decoding attempts by an energy detection test, to discard PUSCH occasions that do not contain transmissions. From this point of view, the energy spent in the preamble could be saved, greatly improving the efficiency of the protocol. However, the possibility of triggering the four-step collision resolution procedure requires the identification of user transmissions, which is made possible by the use of the preambles, which justified their use in the two-step random access protocol [82]. As the channel load increases, the minimum SNR required to attain the target PUPE grows quickly. When plain TIN decoding is applied, the system struggles to support more than $K_a = 8$ active users. Under TIN-SIC decoding, the situation improves. However, at moderate-large numbers of active users (e.g., $K_a = 30$) the scheme operates at more than 8 dB from the achievability bound. The result can be explained by the limited ability of the LDPC code to decode correct messages in the presence of multiple collisions, and by the low number of available PUSCH occasions (at most 64, according to the maximum number of preambles).

Although the results on Gaussian MAC provide a first understanding of the limitations of the two-step random access

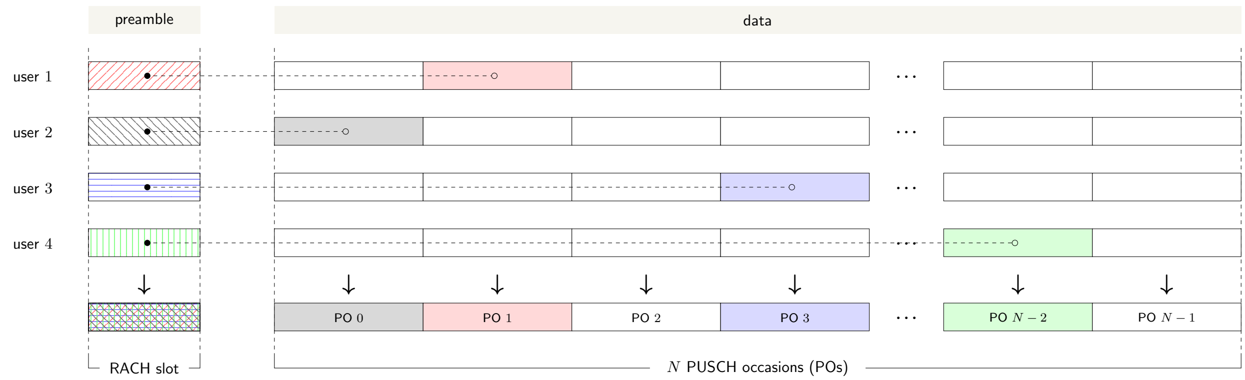


Fig. 7. Illustration of the two-step random access protocol of the 5GNR standard (first transmission only). The number of PUSCH occasions is N.

protocol, the real test bench is transmission over fading channels. Figure 8 reports the performance on the quasi-static Rayleigh channel. The configuration used over the Gaussian MAC is preserved, with the only addition of a pilot field appended to each packet, which is used at the receiver end to estimate the fading channel coefficient (assumed to be independent between users). Therefore, the PUSCH occasions are modified to accommodate 300 channel uses. The overall frame length is $n = 2 \times 139 + 64 \times 300 = 19478$. The target PUPE is set to 10^{-1}. The size of the pilot field amounts to 50 channel uses. The performance (provided only under TIN-SIC) shows again that two-step random access can hardly support large channel loads.

B. Evolution Towards Full Grant-Free Random Access

A simple modification of the scheme, obtained by enlarging the preamble set to 1024 sequences, allows for a remarkable improvement in the supported channel load. Note that the result is obtained while the frame structure remains unchanged, with 64 PUSCH occasions of channel 300 uses each. The reason for the improvement lies in the more accurate channel estimation indirectly provided by the enlarged preamble family. Each preamble points to a PUSCH occasion *and* to a specific pilot sequence: When two users pick different preambles that point to the same PUSCH occasion, by employing two distinct pilot sequences, it is possible to extract sufficiently accurate estimates of the two channel coefficients. This allows to decode with high probability colliding transmissions. A small number of preambles (as in the standard two-step random access scheme) hinders the possibility of having a sufficient diversity of preamble sequences, thus reducing the MPR capability of the receiver. Interestingly, in [3] it was observed that very limited gains could be appreciated by increasing the size of the preamble family beyond 1024. The result could be explained by an additional shortcoming of two-step random access, that is, the limited amount of access patterns (64) allowed by the construction. Inspired by the sparse IDMA scheme of [64], a modification of the two-step random access protocol that relies on packet repetition is outlined in [3]. With each packet repeated ρ times and assuming again $N = 64$ PUSCH occasions, up to $\binom{N}{\rho}$ access patterns can be defined. This allows to make a better use of a larger preamble sets. The performance of this scheme, referred to as sparse block interleave division multiple access (SB-IDMA), is provided in 8. The results are reported assuming either the 5GNR $(500, 100)$ LDPC code [79] or the 5GNR $(500, 100)$ polar code [83] (the dimension of the code refers to the dimension of the outer 11-bits CRC code). The polar code is decoded via the successive cancellation list (SCL) algorithm [84], with a list size set to 128. Large gains can be observed with respect to two-step random access, with a saturation of the performance at high SNR that is mainly driven by the preamble misdetection probability. Additional gains can be noticed by tuning the SB-IDMA scheme parameters: The result is achieved by increasing the length of the preamble to 1778 channel uses, enlarging the preamble set to 8192 preambles, while simultaneously reducing the transmission power of the preamble by a factor $1/12$ and the number of PUSCH occasions to 59 (keeping the overall frame size fixed to 19478). With this dimensioning of the scheme, large channel loads can be supported at a SNR that is only a few tenths of a dB larger than the one required by single-user transmission. The results also highlight the performance improvement that can be achieved by adopting polar codes to protect the data packets. This fact stems from the superior performance of polar codes (with an outer CRC code and under SCL decoding) with respect to LDPC codes, in the short blocklength regime. We expect that the advantage in using polar codes may diminish when transmitting larger packets.

The analysis pinpoints two main limitations of the message A transmission phase of the 5GNR two-step random access protocol, namely (i) the small number of available preamble sequences and (ii) the limited number of access patterns that can be realized. Therefore, it is conceivable that evolutions of 3GPP standards aiming at massive random access scenarios should target the introduction of larger preamble families, as well as richer configurations of access patterns. A fundamental question is how the preamble family should be designed, especially considering the important role that the structure of Zadoff-Chu sequences plays in handling the different propagation delays of user terminals [85, Chapter 11].

VI. Conclusions

The recent introduction of the unsourced multiple access (UMAC) perspective has originated a wave of new massive

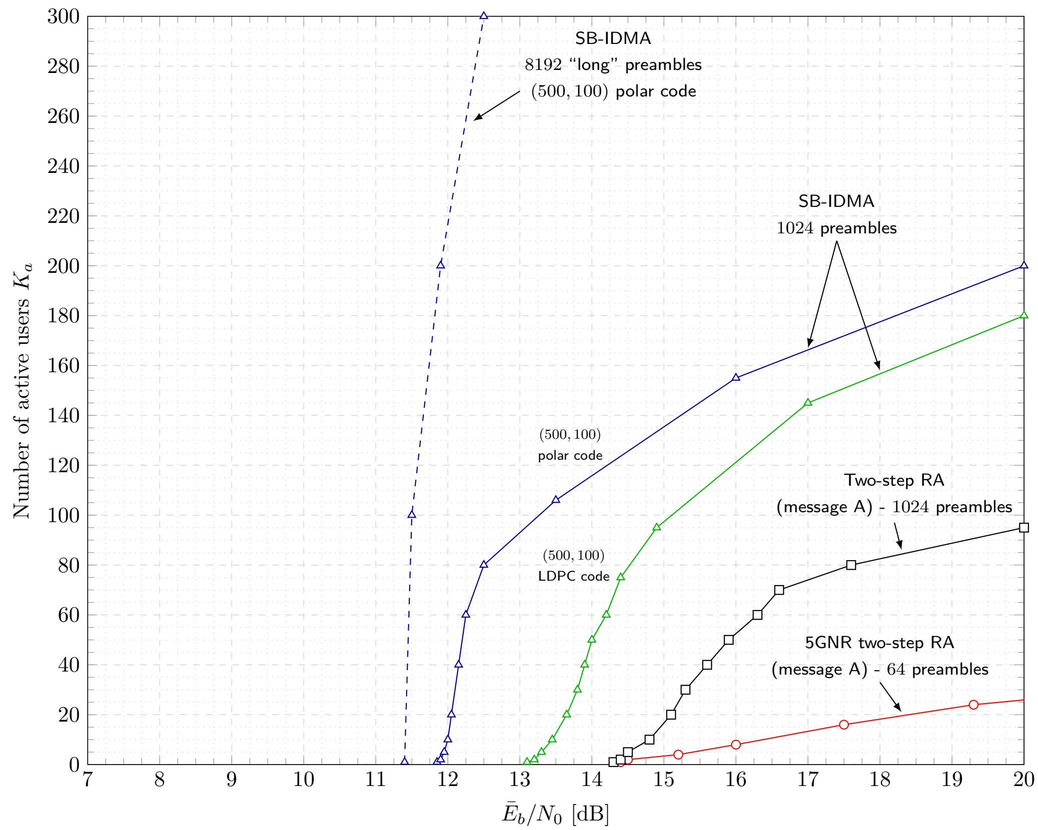


Fig. 8. Minimum average SNR required to achieve a PUPE $= 10^{-1}$, over the quasi-static Rayleigh fading MAC channel. The frame length is $n = 19478$ complex channel uses.

random access protocols, revolutionizing the landscape of large-scale uncoordinated multiple access communications. While the ripples of this wave settle, the crucial question of which UMAC coding architecture will emerge as the reference in future cellular wireless network standards remains open. Scalability, flexibility, and robustness with respect to a broad range of channel conditions and user terminal limitations will play a major role in defining the solution, calling for strict cooperation of coding and information theorists, wireless communication engineers, and hardware architecture designers, in a coordinated effort to refine state-of-the-art UMAC protocol designs.

Acknowledgments

The authors would like to thank Kirill Andreev, Jean-Francois Chamberland, Alexander Fengler, Alexey Frolov, and Khrishna Narayanan for providing numerical results. The authors would also like to thank Maxime Guillaud for the insightful discussions on the tensor-based modulation of [72] and Bobak Nazer for detailed reading of the earlier draft.

References

[1] N. Abramson, "The ALOHA System - Another Alternative for Computer Communications," in *Proc. 1970 Fall Joint Computer Conference*, vol. 37, 1970, pp. 281–285.

[2] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, 2017.

[3] P. Agostini, J.-F. Chamberland, F. Clazzer, J. Dommels, G. Liva, A. Munari, K. Narayanan, Y. Polyanskiy, , S. Stanczak, and Z. Utkovski, "Evolution of the 5GNR Two-Step Random Access towards 6G Unsourced MAC," Apr. 2024, Report V.1.0. [Online]. Available: https://arxiv.org/abs/2405.03348

[4] *5GNR: Medium Access Control (MAC) protocol specification*, 3rd Generation Partnership Project (3GPP) Std. TS 138.321, Rev. 16.1.0, Jul. 2020.

[5] *LoRaWAN Specification v. 1.0*, LoRa Alliance Std., Jan. 2015.

[6] *Short Range Devices; Low Throughput Networks (LTN); Protocols for Radio Interface A*, TS 103 357 V1.1.1, ETSI Std., Jun. 2018.

[7] *ZigBee Specification*, ZigBee Standards Organization Std., Sep. 2012.

[8] D. P. Bertsekas and R. G. Gallager, *Data networks*. Pearson, 2nd Edition, 1991.

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

[10] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part i-carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Trans. Inf. Theory*, vol. 23, no. 12, pp. 1400–1416, Dec. 1975.

[11] B. S. Tsybakov and V. A. Mikhailov, "Free synchronous packet access in a broadcast channel with feedback," *Prob. Peredachi Inform.*, vol. 14, no. 4, pp. 32–59, 1978.

[12] J. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 505–515, Sep. 1979.

[13] J. Massey and P. Mathys, "The collision channel without feedback," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 192–204, Mar. 1985.

[14] L. A. Bassalygo and M. S. Pinsker, "Restricted asynchronous multiple access," *Prob. Peredachi Inform.*, vol. 19, no. 4, pp. 92–96, 1983.

[15] B. S. Tsybakov and A. Rubinov, "Some constructions of conflict-avoiding codes," *Prob. Peredachi Inform.*, vol. 38, no. 4, pp. 268–279, 2002.

[16] S. Ghez, S. Verdu, and S. Schwartz, "Stability properties of slotted Aloha with multipacket reception capability," *IEEE Trans. Autom. Control*, vol. 33, no. 7, p. 640–649, Jul. 1988.

[17] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. COM-25, no. 1, pp. 117–128, January 1977.

[18] S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 32, no. 1, pp. 85–96, 1986.

[19] ——, *Multiuser Detection*. Cambridge University Press, Aug. 1998.

[20] D. Raychaudhuri, "Performance analysis of random access packet-switched code division multiple access systems," *IEEE Trans. Commun.*, vol. 29, no. 6, pp. 895–901, Jun. 1981.

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

[22] A. Polydoros and J. Silvester, "Slotted random access spread-spectrum networks: An analytical framework," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 6, pp. 989–1002, Jul. 1987.

[23] N. Abramson, "VSAT data networks," *Proc. IEEE*, vol. 78, no. 7, pp. 1267–1274, Jul. 1990.

[24] E. Casini, R. D. Gaudenzi, and O. del Rio Herrero, "Contention Resolution Diversity Slotted ALOHA (CRDSA): An Enhanced Random Access Scheme for Satellite Access Packet Networks." *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1408–1419, Apr. 2007.

[25] *Interaction Channel for Satellite Distribution Systems (DVB-RCS)*, EN 301 790 v1.5.1, ETSI Std., May 2009.

[26] Y. Yu and G. B. Giannakis, "High-Throughput Random Access Using Successive Interference Cancellation in a Tree Algorithm," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4628–4639, Dec. 2007.

[27] R. Gallager, "Low-density parity-check codes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 1963.

[28] G. Liva, "Graph-Based Analysis and Optimization of Contention Resolution Diversity Slotted ALOHA," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 477–487, Feb. 2011.

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

[30] K. R. Narayanan and H. D. Pfister, "Iterative collision resolution for slotted ALOHA: An optimal uncoordinated transmission policy," in *Proc. Int. Symp. Turbo Codes and Iterative Inf. Process.*, Gothenburg, Sweden, Aug. 2012.

[31] C. Stefanovic, P. Popovski, and D. Vokobratovic, "Frameless ALOHA Protocol for Wireless Networks," *IEEE Commun. Lett.*, vol. 16, no. 12, pp. 2087–2090, Dec. 2012.

[32] C. Stefanovic and P. Popovski, "ALOHA Random Access that Operates as a Rateless Code," *IEEE Trans. Commun.*, vol. 61, no. 11, Nov. 2013.

[33] E. Sandgren, A. Graell i Amat, and F. Brännström, "On frame asynchronous coded slotted ALOHA: Asymptotic, finite length, and delay analysis," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 691–703, Feb. 2017.

[34] R. De Gaudenzi, O. del Río Herrero, G. Acar, and E. Garrido Barrabés, "Asynchronous Contention Resolution Diversity ALOHA: Making CRDSA Truly Asynchronous," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6193–6206, Nov. 2014.

[35] F. Clazzer, C. Kissling, and M. Marchese, "Enhancing contention resolution ALOHA using combining techniques," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2576–2587, Jun. 2018.

[36] O. Del Rio Herrero and R. De Gaudenzi, "A High Efficiency Scheme for Quasi-Real-Time Satellite Mobile Messaging Systems," in *Proc. Int. Workshop Signal Processing for Space Commun.*, Oct. 2008.

[37] *TS 102 721-1 V1.1.1, Satellite Earth Stations and Systems; Air Interface for S-band Mobile Interactive Multimedia (S-MIM)*, ETSI Std., 2011.

[38] S. Scalise, C. P. Niebla, R. De Gaudenzi, O. Del Rio Herrero, D. Finocchiaro, and A. Arcidiacono, "S-MIM: a novel radio interface for efficient messaging services over satellite," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 119–125, Mar. 2013.

[39] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. John Wiley & Sons, 2006.

[40] Y. Polyanskiy and Y. Wu, *Information theory: From coding to learning*. Cambridge University Press, 2024.

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

[42] Y. Polyanskiy, "Information theoretic perspective on massive multiple-access," Tutorial at 2018 North-American School of Information Theory, Jun. 2018.

[43] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017.

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

[45] E. Marshakov, G. Balitskiy, K. Andreev, and A. Frolov, "A Polar Code Based Unsourced Random Access for the Gaussian MAC," in *Proc. IEEE Vehicular Technology Conference Fall*, Honolulu, HI, USA, Sep. 2019.

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

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

[48] K. Andreev, E. Marshakov, and A. Frolov, "A Polar Code Based TIN-SIC Scheme for the Unsourced Random Access in the Quasi-Static Fading MAC," in *Proc. IEEE Int. Symp. Inf. Theory*, Los Angeles, CA, USA, Jun. 2020.

[49] E. De Carvalho, E. Bjornson, J. H. Sorensen, E. G. Larsson, and P. Popovski, "Random Pilot and Data Access in Massive MIMO for Machine-Type Communications," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 7703–7717, Dec. 2017.

[50] J. H. Sorensen, E. De Carvalho, C. Stefanovic, and P. Popovski, "Coded Pilot Random Access for Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8035–8046, Dec. 2018.

[51] M. J. Ahmadi, M. Kazemi, and T. M. Duman, "Unsourced Random Access Using Multiple Stages of Orthogonal Pilots: MIMO and Single-Antenna Structures," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 1343–1355, Feb. 2024.

[52] M. Ozates, M. Kazemi, and T. M. Duman, "A Slotted Pilot-Based Unsourced Random Access Scheme with a Multiple-Antenna Receiver," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3437–3449, Apr. 2024.

[53] A. Fengler, O. Musa, P. Jung, and G. Caire, "Pilot-Based Unsourced Random Access With a Massive MIMO Receiver, Interference Cancellation, and Power Control," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 5, pp. 1522–1534, May 2022.

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

[55] R. Calderbank and A. Thompson, "CHIRRUP: a practical algorithm for unsourced multiple access," *Information and Inference: A Journal of the IMA*, vol. 9, no. 4, p. 875–897, Dec. 2019.

[56] S.-C. Chang and J. Wolf, "On the T-user M-frequency noiseless multiple-access channel with and without intensity information," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 41–48, Jan. 1981.

[57] A. Fengler, P. Jung, and G. Caire, "SPARCs for unsourced random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6894–6915, Oct. 2021.

[58] A. Joseph and A. R. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2541–2557, May 2012.

[59] R. Venkataramanan, S. Tatikonda, and A. Barron, "Sparse regression codes," *Found. Trends® Commun. Inf. Theory*, vol. 15, no. 1-2, pp. 1–195, 2019.

[60] J. R. Ebert, V. K. Amalladinne, J.-F. Chamberland, and K. R. Narayanan, "A hybrid approach to coded compressed sensing where coupling takes place via the outer code," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Toronto, ON, Canada, Jun. 2021.

[61] V. K. Amalladinne, A. K. Pradhan, C. Rush, J.-F. Chamberland, and K. R. Narayanan, "Unsourced random access with coded compressed sensing: Integrating AMP and belief propagation," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2384–2409, Apr. 2021.

[62] J. R. Ebert, V. K. Amalladinne, S. Rini, J.-F. Chamberland, and K. R. Narayanan, "Coded Demixing for Unsourced Random Access," *IEEE Trans. Signal Process.*, vol. 70, pp. 2972–2984, Jun. 2022.

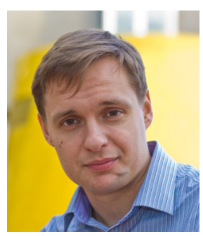
[63] P. Agostini, Z. Utkovski, and S. Stanczak, "BiSPARCs for Unsourced Random Access in Massive MIMO," in *Proc. 57th Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, US, Nov. 2023.

[64] A. K. Pradhan, V. K. Amalladinne, A. Vem, K. R. Narayanan, and J.-F. Chamberland, "Sparse IDMA: A Joint Graph-Based Coding Scheme for Unsourced Random Access," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7124–7133, Nov. 2022.

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

[66] D. Truhachev, M. Bashir, A. Karami, and E. Nassaji, "Low-complexity coding and spreading for unsourced random access," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 774–778, Mar. 2021.
[67] E. Nassaji, M. Bashir, and D. Truhachev, "Unsourced Random Access Over Fading Channels via Data Repetition, Permutation, and Scrambling," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1029–1042, Feb. 2022.
[68] J. Yan, G. Song, Y. Li, and J. Wang, "ODMA Transmission and Joint Pattern and Data Recovery for Unsourced Multiple Access," *IEEE Wireless Commun. Lett.*, vol. 12, no. 7, pp. 1224–1228, Jul. 2023.
[69] M. Ozates, M. Kazemi, and T. M. Duman, "Unsourced Random Access Using ODMA and Polar Codes," *IEEE Wireless Commun. Lett.*, vol. 13, no. 4, pp. 1044–1047, Apr. 2024.
[70] R. Schiavone, G. Liva, and R. Garello, "Design and Performance of Enhanced Spread Spectrum Aloha for Unsourced Multiple Access," *IEEE Commun. Lett.*, 2024, accepted.
[71] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan, and J.-F. Chamberland, "Polar coding and random spreading for unsourced multiple access," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, Jun. 2020.
[72] A. Decurninge, I. Land, and M. Guillaud, "Tensor-Based Modulation for Unsourced Massive Random Access," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 552–556, Mar. 2021.
[73] Z. Han, X. Yuan, C. Xu, S. Jiang, and X. Wang, "Sparse Kronecker-Product Coding for Unsourced Multiple Access," *IEEE Wireless Commun. Lett.*, vol. 10, no. 10, pp. 2274–2278, Oct. 2021.
[74] M. Gkagkos, K. R. Narayanan, J.-F. Chamberland, and C. N. Georghiades, "FASURA: A Scheme for Quasi-Static Fading Unsourced Random Access Channels," *IEEE Trans. Commun.*, vol. 71, no. 11, pp. 6391–6401, Nov. 2023.
[75] *LTE: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN)*, 3rd Generation Partnership Project (3GPP) Std. TS 136.300, Rev. 10.5.0, Nov. 2011.
[76] *LTE: Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification*, 3rd Generation Partnership Project (3GPP) Std. TS 136.300, Rev. 13.2.0, Aug. 2016.
[77] *5GNR: Medium Access Control (MAC) protocol specification*, 3rd Generation Partnership Project (3GPP) Std. TS 138.321, Rev. 15.3.0, Sep. 2018.
[78] E. Dahlman, S. Parkvall, and J. Sköld, *5GNR The Next Generation Wireless Technology*. Academic Press, 2018.
[79] T. Richardson and S. Kudekar, "Design of low-density parity check codes for 5G new radio," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 28–34, Mar. 2018.
[80] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, US, Nov. 1993.
[81] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[82] C.-C. Tseng, H.-C. Wang, J.-R. Chang, L.-H. Wang, and F.-C. Kuo, "Design of Two-Step Random Access Procedure for URLLC Applications," *Wireless Personal Communications*, vol. 121, no. 2, pp. 1187–1219, Nov. 2021.
[83] V. Bioglio, C. Condo, and I. Land, "Design of Polar Codes in 5G New Radio," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 29–40, Jan. 2020.
[84] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.
[85] E. Dahlman, S. Parkvall, and J. Sköld, *4G LTE-Advanced Pro and The Road to 5G*. Academic Press, 2016.

Gianluigi Liva (M'08–SM'14) was born in Spilimbergo, Italy, in 1977. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Bologna, Italy, in 2002 and 2006, respectively. From 2004 to 2005, he was visiting at the University of Arizona, Tucson. Since 2006, he has been with the Institute of Communications and Navigation, German Aerospace Center (DLR), where he currently leads the Information Transmission Group. Since 2014, he has been a Lecturer of channel codes with iterative decoding with LNT, TUM. He received the Italian National Scientific Habilitation (ASN) as Full Professor in Telecommunication Engineering in July 2017. He is/has been active in the DVB-RCS and DVB-S2 standardization groups, and in the standardization of error correcting codes for space communications within the CCSDS. Since 2020, he serves as Associate Editor in Coding and Information Theory for the IEEE Transactions on Communications.



Yury Polyanskiy (S'08-M'10-SM'14-F'24) is a Cutten Professor of Electrical Engineering and Computer Science, a member of Institute for Data, Systems, and Society (IDSS) and Laboratory for Information and Decision Systems (LIDS) at MIT. Yury received M.S. degree in applied mathematics and physics from the Moscow Institute of Physics and Technology, Moscow, Russia in 2005 and Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ in 2010. His research interests span information theory, machine learning and statistics. Dr. Polyanskiy won the 2020 IEEE Information Theory Society James Massey Award, 2013 NSF CAREER award and 2011 IEEE Information Theory Society Paper Award.