TOWARD 5G WIRELESS INTERFACE TECHNOLOGY

Fan Wei, Wen Chen, Yongpeng Wu, Jun Li, and Yuan Luo

Designers of fifth-generation (5G) wireless networks are focusing on enabling highly reliable low-latency communications that support a high data rate and allow massive connectivity. Nonorthogonal multiple access (NOMA), an essential enabling technology tailored to accommodate a wide range of communication requirements, shows potential as a tool for helping 5G networks to fulfill these promised capabilities. By coordinating connections for massive numbers of devices within the same resource block on power domain, frequency domain, or code domain, NOMA is superior to conventional orthogonal multiple access in terms of network connectivity, system throughputs, and other characteristics. Sparse code multiple access (SCMA) is a kind of multicarrier code-domain NOMA and has been studied extensively. The challenge in designing a high-quality SCMA system is crafting feasible encoding and decoding schemes to meet the desired requirements. In this article, we describe recent progress in designing multidimensional codebooks, a practical low-complexity decoder, and grant-free multiple access for SCMA systems. Our particular focus is on showing how the designs of the multidimensional constellation and factor graphs (FGs) have formed the foundation of SCMA codebooks. In addition, we review various low-complexity SCMA decoders with a special emphasis on sphere decoding. Also, we introduce SCMA grant-free transmission based on the framework of belief propagation (BP) and discuss the problem of collision resolution.

Digital Object Identifier 10.1109/MVT.2018.2867280
Date of publication: 1 October 2018

Background

New features for the next generation of wireless networks have emerged from the 3rd Generation Partnership Project. These features include enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultrareliable low-latency communication (URLLC). The aim of eMBB is to permit a high data rate for wireless networks because the typical 5G transmission rate will be 1 Gb/s, with the peak rate faster than 10 Gb/s, which is roughly ten or 100 times faster than that of the fourth-generation networks. Besides enabling these speeds, mMTC should accommodate massive connectivity by allowing networks to connect millions of devices within a single square kilometer, a capacity far greater than that permitted by today's conventional networks. Meanwhile, URLLC should reduce end-to-end delay to a matter of milliseconds, thereby supporting such communication capabilities as real-time mobile control and vehicle-to-vehicle applications and communications.

Conventional orthogonal multiple access, such as time-division multiple access, code-division multiple access, and orthogonal frequency-division multiple access (OFDMA), assigns the orthogonal resource to each user exclusively. However, as massive numbers of devices are connected at the same time, available time–frequency resources quickly become scarce. With massive connectivity, the inefficiency of giving every user exclusive access to such limited resources becomes apparent. Moreover, channel codes, as they have evolved, have pushed conventional point-to-point transmission nearly to the Shannon limits. Thus, it is essential to explore network communication using next-generation wireless interface technologies that take full advantage of the spectral efficiency, offer more throughputs, and accommodate more connected users. To this end, NOMA, which allows multiple users to transmit simultaneously within the shared resources, represents a promising frontier for study.

Developers have proposed various NOMA techniques based on differing principles. These include power-domain NOMA, low-density signatures (LDSs), SCMA, and pattern-division multiple access. While the techniques have different rationales, they share an underlying goal: that multiple users should be able to transmit simultaneously in each orthogonal time–frequency resource. Here, we comprehensively review recent progress on SCMA [1], a code-domain NOMA. The focus is on designing codes to make SCMA especially reliable. Also, a practical low-complexity decoder for the multiuser system needs to be developed to accommodate massive connectivity demands. To enable the low-latency communication, grant-free multiple access may be the preferred alternative to conventional request–grant multiple access. Thus, it is interesting to consider SCMA combined with grant-free transmission. We specifically look at the low-complexity grant-free receiver design and the collision resolution problem within the contention-based transmission.

SCMA

In this section, we provide a brief introduction to SCMA. Figure 1 depicts the block diagram for the uplink SCMA system. In the encoder, coded bits are modulated to the multidimensional sparse codewords consisting of the high-dimensional lattice points. As an example, Figure 2 illustrates the four-dimensional sparse codewords with two nonzero entries. The nonzero entries can be modulated to OFDMA subcarriers or the multiple-input, multiple-output (MIMO) antennas. Note that choosing the positions of nonzero entries can be viewed as a combinational problem. Therefore, at most, $\binom{K}{N}$ users can be accommodated, where K is the total number of subcarriers and N denotes the number of nonzero entries. To enable the sparse structure of codewords, the number of nonzero entries should satisfy $N \ll K$. This reduces the likelihood that users will experience collisions in a single subcarrier and further simplifies decoding. For the four-dimensional codewords in Figure 2, the maximum number of users is six, whereas, due to the sparsity of codewords, the number of possible collisions among users is reduced to three per subcarrier.

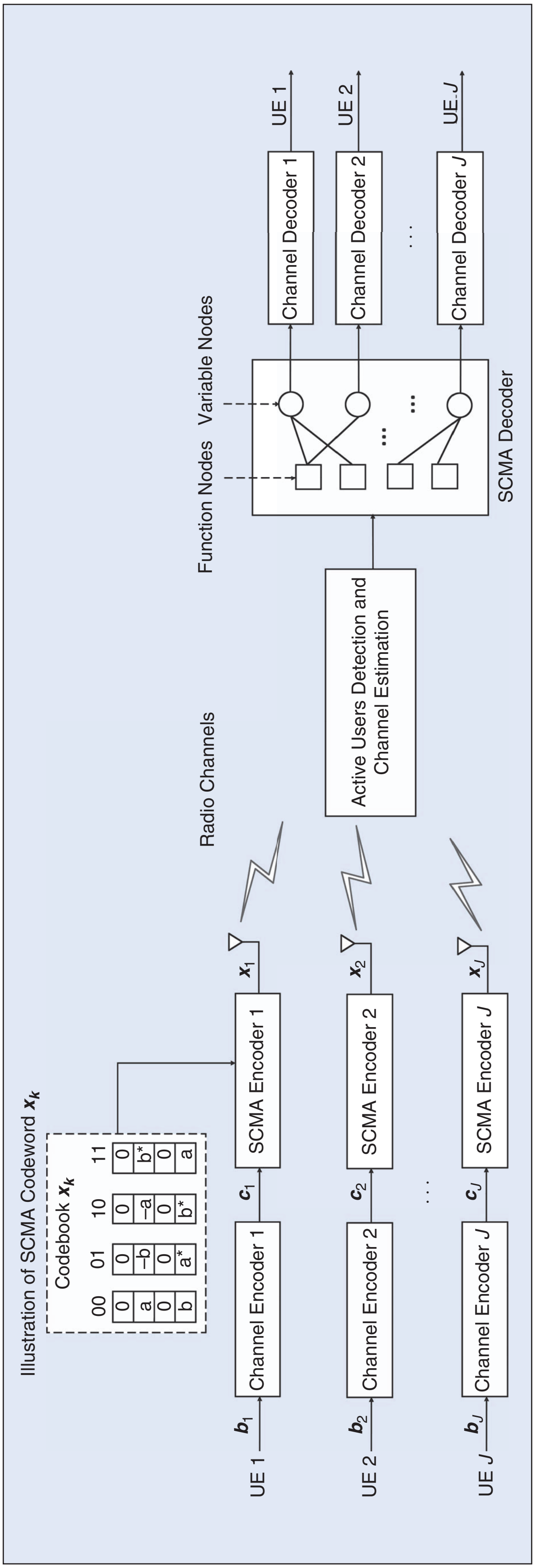


FIGURE 1 A block diagram of SCMA systems.

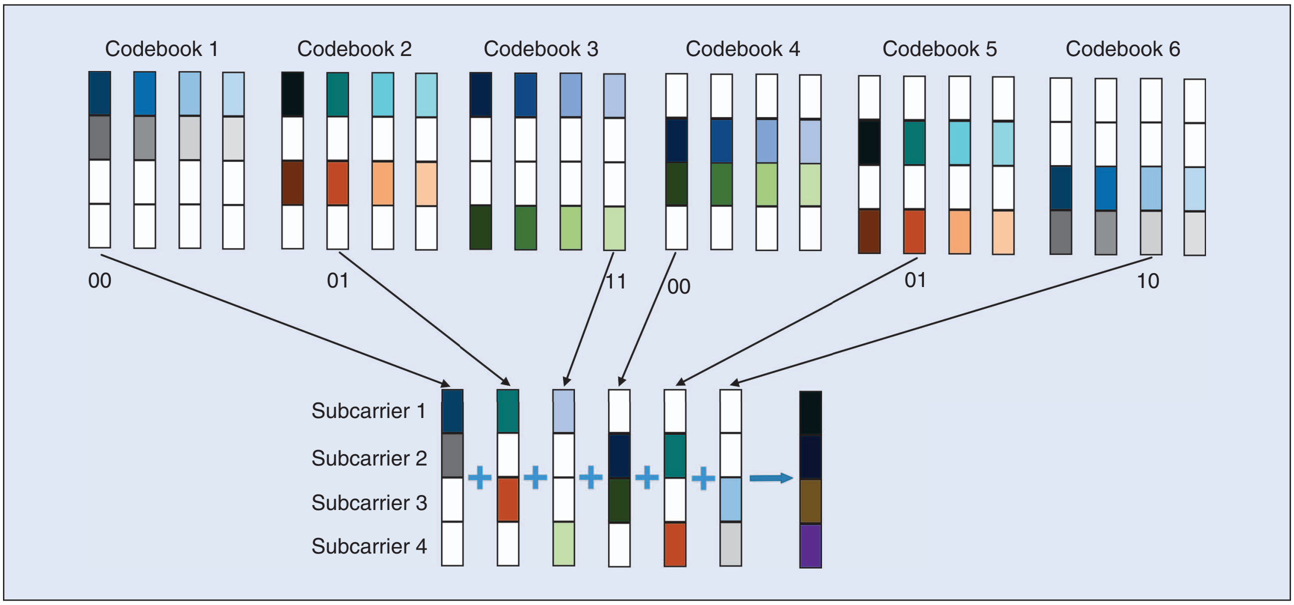


FIGURE 2 A four-dimensional SCMA codebook structure.

The SCMA codebook construction can be formulated as a suboptimal multiple-stage optimization problem. Initially, a mother constellation based on high-dimensional lattice points was constructed. Thereafter, lattice rotation for the mother constellation was further applied to optimize the product distance of the constellation and induce the dependency and power variations among the lattice dimensions. The subsequent design of the codebook addresses the needs of distinct users by introducing the constellation function operators, including complex conjugates, phase rotations, and dimensional permutations for the lattice constellations. Those operators only render the distinct codewords for the users while leaving the minimum Euclidean distance of the constellations unchanged.

After receiving the superimposed codewords, the decoder performs a channel estimation and data decoding based on pilots as well as data signals. The active user identification is also necessary if grant-free transmission is considered. The sparse structure of SCMA codewords can be represented by an FG as in low-density parity check (LDPC) codes. This structure enables the decoder to employ the message-passing algorithm (MPA) for the iterative decoding of the data.

SCMA Encoder Design and the Performance Limits Analysis

As a code-domain NOMA, SCMA requires, on average, less energy for signal constellation because of the shaping gain of multidimensional codewords. Thus, the codebook design helps the SCMA system to perform better. In this section, we review the literature about SCMA codebook design and provide insights regarding performance limits of the SCMA system.

SCMA Codebook Design

As a multicarrier NOMA, the codebooks of SCMA are essentially multidimensional constellation points. SCMA, as originally planned, presented a systematic construction method for codebook design where the high-dimensional constellation points were generated based on square-quadrature amplitude modulation (QAM). Star-QAM has emerged as an alternative approach offering better bit-error-rate (BER) performance in the fading channels. It also outperforms the square-QAM in peak-power-limited systems. Consequently, researchers brought forward a star-QAM-based codebook design for SCMA [2]. By optimizing the coordinates of the signal points in different rings of the star-QAM, the star-QAM-based SCMA enjoys a larger minimum Euclidean distance compared with LDS and the codebooks in [1].

To leverage the ratio of peak-to-average power as well as the spectrum efficiency, spherical codes have been adapted as another approach for the high-dimensional codebook design of SCMA [3]. Various good spherical codes have been constructed through binary codes, shells of lattices, permutations of vectors or computer searching, and by other means. The optimal spherical codes with low dimensions were shown to have large coding gains. An SCMA codebook with good performance can be obtained by concatenating those low-dimensional spherical codes.

Apart from the mother constellation design, the FG is another important feature in elevating the performance of the SCMA system. A properly designed FG may expand system capacity, reduce the interference of each subcarrier, and reduce decoding complexity for the receiver. In [4], the capacity of uplink SCMA is derived when the covariance matrix of the codewords is diagonal. The FG design is formulated as an integer programing optimization problem with the object function being the sum rate of the system. Each link between the variable node and function node in the FG is determined by maximizing the individual user rate in one subcarrier iteratively. In Figure 3, the throughputs of different FG construction and power-allocation schemes are compared. This shows that, through the optimization of FGs, some users may occupy one subcarrier exclusively and experience no interference from other users. In other words, the irregular SCMA was shown to have a better system performance.

Performance Limits of the SCMA System

A system based on multidimensional SCMA codewords can be viewed as equivalent to an MIMO system. In addition, the MPA receiver may reduce or even eliminate the multiuser interference in each subcarrier. Thus, the SCMA system can be treated as a kind of single-user MIMO system. With the known results of MIMO, the process of deriving the lower bound of symbol error rate for SCMA is straightforward [5]. Furthermore, an inspection of the pair-wised probability of MIMO SCMA with a large-scale fading channel shows that the diversity order of SCMA is mainly due to two factors: the number of received antennas and the signal-space diversity of the SCMA multidimensional codewords [6]. In addition, evidence shows that large-scale fading only affects the coding gain of the system.

SCMA Decoder Design

The invention of turbo codes has inspired developers to apply the turbo principle in the design of communication systems. The turbo principle has also led developers to implement soft information between the SCMA decoder and the forward error correction decoder. This technique lowers the coded BER of SCMA [7]. However, improved BER performance increases decoding complexity for the receiver. This is why a low-complexity decoder is especially essential for SCMA.

As the LDPC codes, SCMA can be represented by the FG. Thus, it is natural to use the BP or MPA for the SCMA decoder. Due to the sparsity of the codewords, MPA has reasonable decoding complexity for the low and middle modulation order systems. However, when the number of users is increasing or the modulation order is high (e.g., the 32- or 64-ary SCMA), decoding becomes prohibitively complex. As such, satisfactory low-complexity decoders for SCMA are needed.

One such decoder is based on partial marginalization [8], an idea grounded in the observation that the convergence rate of message passing for different users may be different. Partial marginalization demonstrates that it is better to analyze the transmit symbols for users with faster convergence rates before the algorithm termination. Those symbols can be subtracted from the received signals. For the remaining users, message passing will only be involved in the subsequent iterations. Theoretically, the complexity of partial marginalization is exponential, and the performance is highly dependent on which portion of users are chosen to be judged first.

TO ENABLE THE LOW-LATENCY COMMUNICATION, GRANT-FREE MULTIPLE ACCESS MAY BE THE PREFERRED ALTERNATIVE TO CONVENTIONAL REQUEST–GRANT MULTIPLE ACCESS.

While most work in this area focuses on simplifying decoding on the receiver side, a proper constellation design can also streamline computing demands [9]. The idea is shown in Figure 4. For the multidimensional SCMA codewords with a modulation order M, one can project that to the constellation with $m < M$ points. Evidently, this will cause some overlaps on the point labeling. Nevertheless, the decoder is able to distinguish points from one another as long as the overlap points are separated on the other dimensions (e.g., points 01 and 10 overlap on dimension 1 but are separated on dimension 2). Since $m < M$, the decoding complexity is reduced from M^{dc} to m^{dc}, where d_c is the number of collision users in each subcarrier.

Researchers working on improving MPA are concentrating on updating the function nodes. In each iteration, the MPA decoder seeks the point that is most likely to be transmitted. However, the scheme requires an exhaustive search through all possible combinations of transmitted signals, making it inefficient. As an alternative

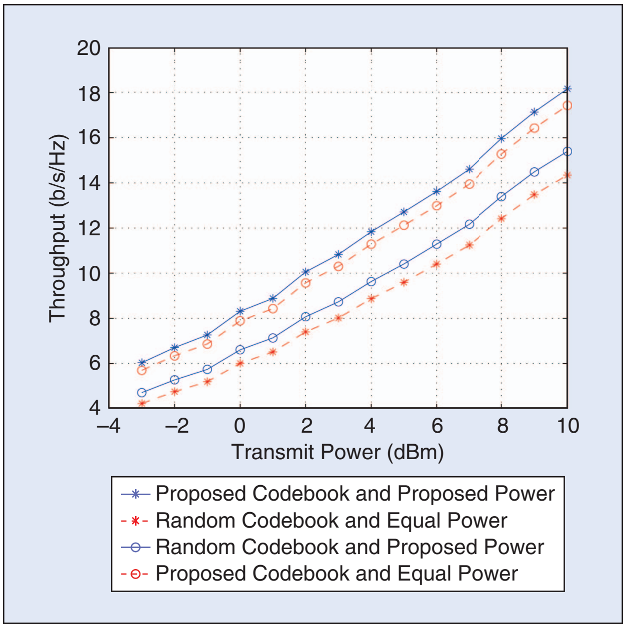


FIGURE 3 Throughput comparisons for different codebook and power allocation schemes.

THE SCMA CODEBOOK CONSTRUCTION CAN BE FORMULATED AS A SUBOPTIMAL MULTIPLE-STAGE OPTIMIZATION PROBLEM.

approach, the decoder can also proceed in a more efficient way by considering only the signal points near the received signals. The idea is essentially the same as that of a sphere decoder. For the SCMA, the decoder seeks a candidate list first by list-sphere decoding (LSD). Next, MPA can be operated within the candidate list, thus simplifying decoding [10], [11]. The size of the candidate set can be tuned to balance the decoding complexity and the performance. Essentially, a sphere decoder can be viewed as a kind of depth-first tree search algorithm.

Figure 5 shows an example of a searching tree with $L = 5$ levels. The search process starts from the root node and proceeds to the leaf nodes. All solid lines correspond to the survival paths during the search, while the dotted lines correspond to the eliminated paths. The decoding complexity depends on the number of visited nodes on the binary tree during the search process. Thus, by avoiding the unnecessary visits to nodes, a simpler search is possible if some branches of the tree are pruned properly. SCMA is a kind of multisubcarrier NOMA. As such, instead of carrying out sphere decoding on different subcarriers independently, the messages are exchanged from different subcarriers so that some branches may be pruned in advance on the lower search levels. More specifically, when a point is reported to be excluded in the list by one subcarrier, the other subcarriers can circumvent the route to this point by pruning the unnecessary branches of the searching tree.

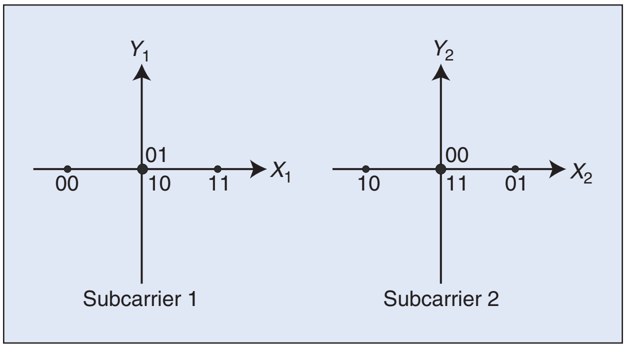


FIGURE 4 A low number of projections for a four-ary SCMA.

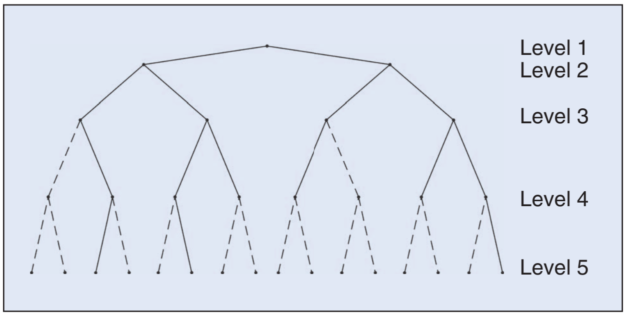


FIGURE 5 A sphere decoder depth-first tree search.

Figure 6(a) shows the number of visited nodes with node pruning (NP-LSD-MPA) versus the conventional methods (LSD-MPA). Figure 6(b) demonstrates that node pruning results in improved BER performance. In fact, an LSD error occurs when the actual transmit point is not contained within the candidate list. Hence, if the redundancy nodes are pruned within the list, the actual transmit point would likely enter the final list and, thus, improve the performance. A sphere decoder for SCMA also has the advantage of being applicable to the low number of projection codewords proposed in [9] (see [11] for a detailed discussion).

SCMA with Grant-Free Transmission

Another challenge for the advanced multiple-access technique developers is to address the issue of user scheduling. In long-term evolution networks, user scheduling is realized through the request-grant procedure, where active users periodically send scheduling requests to the base station (BS) and wait for the grants as well as resource assignments. However, the handshakes between the BS and the active users may result in considerable delays. Meanwhile, the dynamic signaling degrades the spectral efficiency of the network. Thus, contention-based grant-free transmission is advocated in the future 5G network. Instead of the handshakes, active users contend the shared resources and transmit directly to the BS. However, two problems arise in grant-free multiple access: user identification and collision resolution.

Identification of Active Users for Grant-Free SCMA

For typical 5G scenarios, such as those involving mMTC, the number of users connected to the network is presumed to be huge. While millions of devices are connected to the network, only a few users are communicating at the same time. Thus, the task of identifying active users is generally framed as a kind of compressed sensing problem.

With the understanding that inactive users are sending no data symbols and have no channel responses, the channels from all users can be viewed as a sparse vector with nonzero entries representing active users. Therefore, active users can be recognized by detecting support of the sparse vector. In [12], three algorithms are proposed that use pilot signals only to estimate the sparse channel vectors.

Generally, channel estimates based on data are more accurate than those based only on pilot signals. Further, user activity information is contained not only in the received pilot signals but also the data signals. Thus, one proposed receiver would, by exploiting both pilot and data signals, perform joint user identification, channel estimation, and data decoding [13]. For the problem of estimating hybrid continuous (channel coefficients) and discrete (data symbols) variables, direct maximum likelihood detection is unrealistic because it is so complex. Thus, a low-complexity iterative decoder based on the FG of SCMA is necessary.

The FG representation of SCMA with estimation of hybrid variables is formulated in Figure 7(a). The representation shows three loops: one for estimating channels, one for detecting data, and one for detecting user activity. In the data-detection loop, the term $p(\boldsymbol{x}_k | c_k)$ denotes the mapping function of SCMA encoder k, which maps the coded bits to the SCMA codewords. The function node f_{tn} corresponds to the likelihood function on the tth time slot, subcarrier n. For the channel-estimation loop, the tapped-delay channel model with length L is considered. The function $\phi_{n,k}$ corresponds to the Fourier transformation of channel responses from the lag domain to the frequency domain. To facilitate the detection of sparse signals, the two-layer hierarchy channel model is used for modeling each channel tap h_{kl}. On the first layer, the channel is assumed to follow the Gaussian distribution with variance λ_{kl}. On the second layer, the variance is further modeled as a gamma distribution $\Gamma(\lambda_{kl} | a, b)$ with a noninformative priori parameter of, for instance, $a = 10^{-7}$ and $b = 10^{-7}$. Such a formulation of channel responses leads to the Student's t distribution when integrating out the variables λ_{kl}. The Student's t distribution exhibits heavy tails and thus favors sparse solutions for the problem of identifying active users.

Essentially, a sphere decoder can be viewed as a kind of depth-first tree search algorithm.

With the formulation of FG, the message passing for data symbols may be calculated based on BP. However, the direct BP updating is cumbersome for the hybrid continuous and discrete variable models. The idea for rendering a tractable computation is to approximate the virtual variables, defined as the production of discrete data symbols and channel coefficients, into some continuous distributions for the sake of low-complexity computation; see Figure 7(b) for illustrations. To ensure the accuracy of approximation, the true distribution of variables is projected into Gaussian distribution for each user individually with the minimized Kullback–Leibler (KL) divergence. This coincides with the idea of expectation propagation (EP) message passing. With Gaussian distributed messages, BP updating is now tractable by some simpler computation.

Figure 8, which demonstrates the performance of a receiver based on EP message passing, examines machine-type communications where 256 coded bits are contained in each packet. The user identification comparison by the joint and pilot-aided-only detector is shown in Figure 8(a). The activity for each user is determined by the power of the estimated channels since inactive users experience the equivalent of zero channel coefficients. Clearly, by exploring the pilot and data symbols concurrently, the joint detector can achieve a much lower error-detection rate. Figure 8(b)

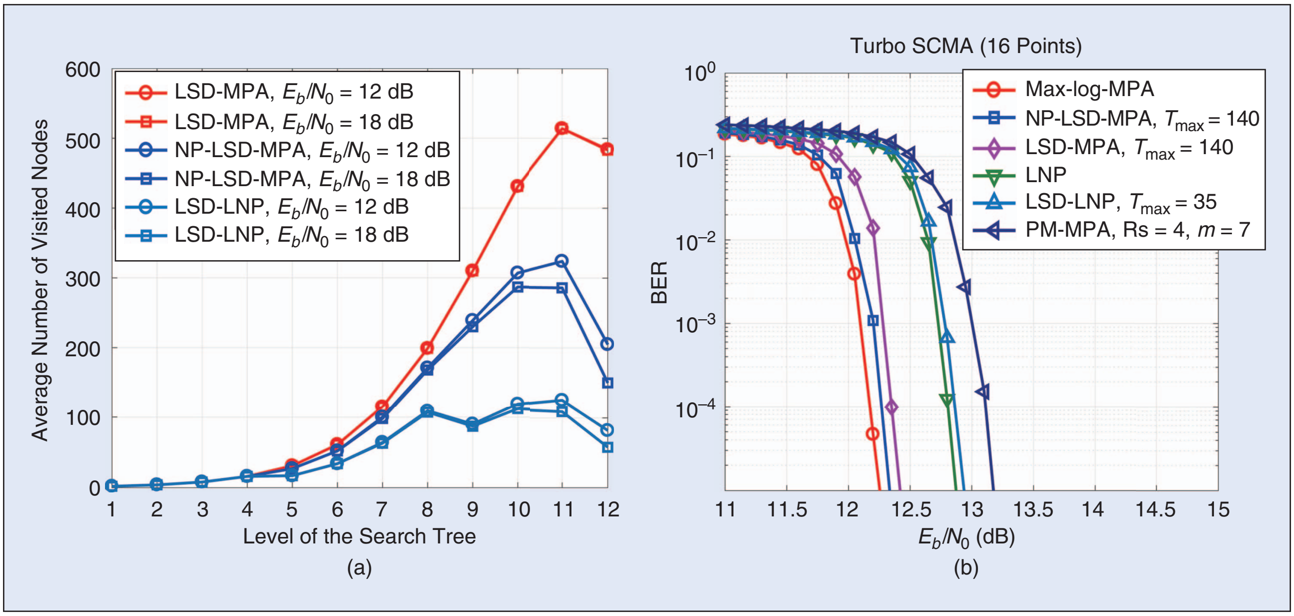


FIGURE 6 An evaluation of (a) the average number of nodes visited in each level of the searching tree for the 150% overloaded 16-ary SCMA system and (b) the BER performance of the 150% overloaded 16-ary SCMA system. LSD-MPA corresponds to the list sphere decoder. NP-LSD-MPA denotes the LSD with node pruning, and LNP is the low number of projected SCMA codewords. The conventional max-log-MPA is used as a performance benchmark. PM-MPA denotes the decoder based on partial marginalization [8].

compares the BER performance for different decoding algorithms. The pilot-aided MPA (P-MPA) algorithm, which relies on pilots only for channel estimation and user detection, has the worst performance. Other joint message-passing receivers, such as BP-MF and BP-GA, are less effective due to their inaccurate approximation for the true probability. Because the probability estimate stems from minimized KL divergence, the performance gap between the BP-GA-EP receiver [13] and a receiver (Genie) with exact channel state information and user activity is only roughly 1 dB.

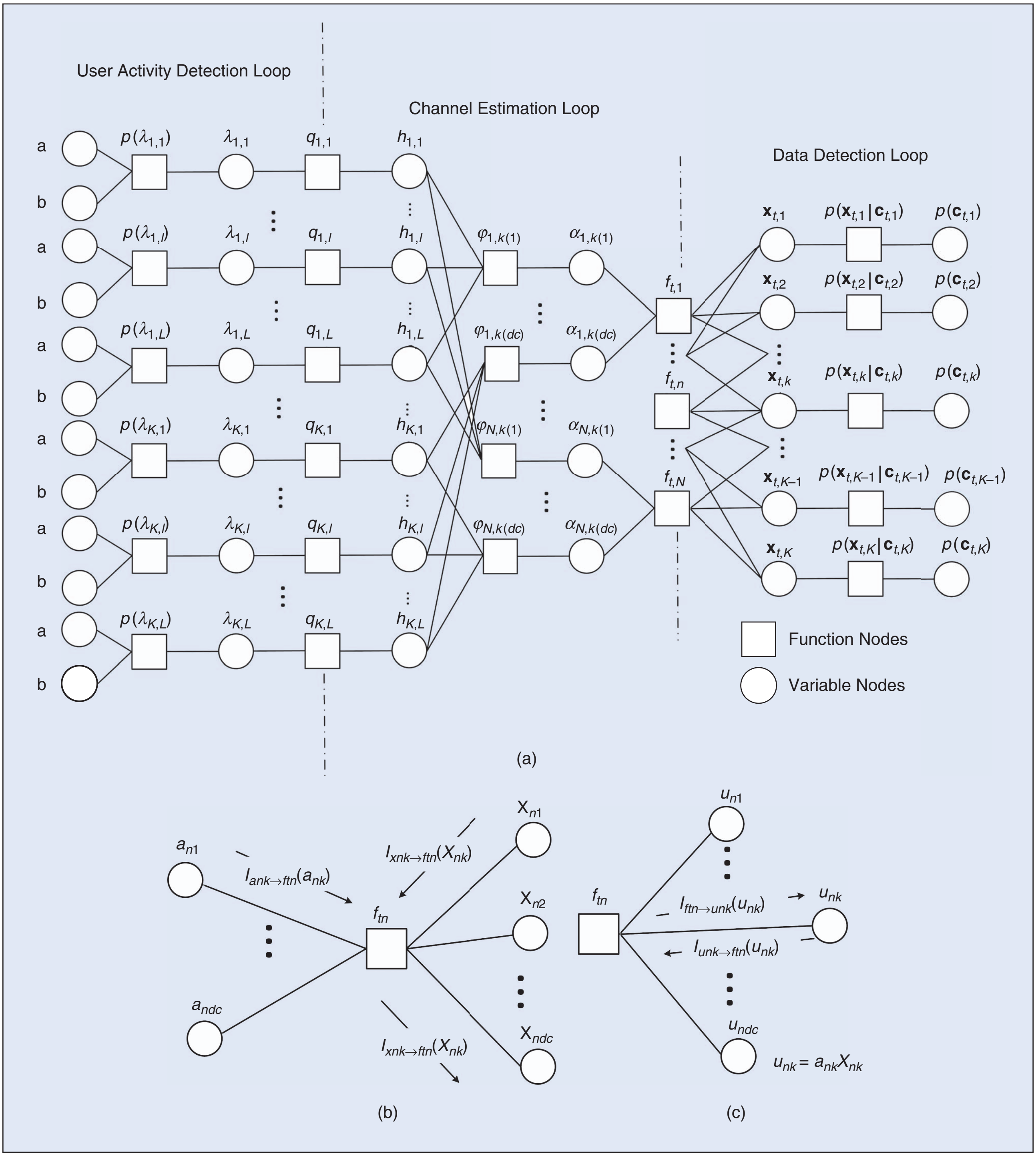


FIGURE 7 The FG representation of (a) the SCMA with a hybrid variables estimation and (b) and (c) the message passing within function node *n*. (b) The diagram shows the message passing with direct BP. (c) The diagram shows the message passing with EP approximation, i.e., the extrinsic messages for virtual variable *u*, which is the production of channel *h* and data *x*, are approximated with Gaussian distributions with the minimized KL divergence.

Collision Resolution for Grant-Free SCMA

The radio resources for grant-free transmission are defined in [12]. The basic radio resource, which is referred to as a *contention transmission unit* (*CTU*), is defined as the combination of time data, frequency data, SCMA codebook data, and pilot data. The user equipment (UE) is (possibly) mapped to CTUs via the mapping rule $CTU_{\text{index}} = UE_{\text{ID}}$ mod N_{CTU}, where N_{CTU} denotes the number of CTUs. In mMTC, the number of UEs would always exceed the number of available CTUs. As a consequence, multiple UEs can be mapped to the same CTU for data transmission. A collision happens when two or more active UEs transmit concurrently within the same CTU.

To resolve the UE collision problems, the conventional approach conducts a retransmission by the random back-off mechanism. However, if the static mapping rule is used, constant collisions may happen when bursty traffic occurs even when the random back-off is used. Further, some CTUs may be allocated with more UEs due to the number of UEs exceeding that of CTUs. To circumvent the prior described issues, an UE-to-CTU mapping rule based on acknowledgment (ACK) feedback is proposed [14]. The idea is to use the conventional static mapping rule in the first round of handshaking between BSs and UEs. Next, an ACK list containing the detection and decoding results for all UEs is fed back from the BSs to receivers. For each UE, two bits are assigned to indicate the status: (0, 0) for no data received, (1, 1) for data successfully decoded, and (0, 1) for retransmission needed. Figure 9 is an illustration of the mapping rule based on ACK feedback.

Figure 10 depicts a detailed grant-free transmission procedure with a mapping rule based on ACK feedback. Transmission delay is assumed to be three time slots, and MAPi denotes the mapping rule generated in the *i*th time slot. To start with, MAP1 in the first time slot is set based on a conventional static mapping rule. Afterwards, with the receipt of ACK from the BS, an adaptive MAP4 is determined by UEs. Specifically, for UE j, if the code is 01, indicating that UE j needs to retransmit, an exclusive CTU is immediately assigned again. The BS counts the number of CTUs associated with only one UE, i.e., N_{single} and set $CTU_{\text{index}} = j \bmod N_{\text{single}}$. If the code is not 01, an exclusive CTU is unnecessary. The BS now counts the number of CTUs associated with multiple UEs, i.e., N_{multi} and set $CTU_{\text{index}} = j \bmod N_{\text{multi}} + N_{\text{single}}$. The subsequent slots can be handled in a similar manner. Theoretical analysis of methods based on ACK feedback shows that the collision probability would always be less than that of conventional static mapping rules.

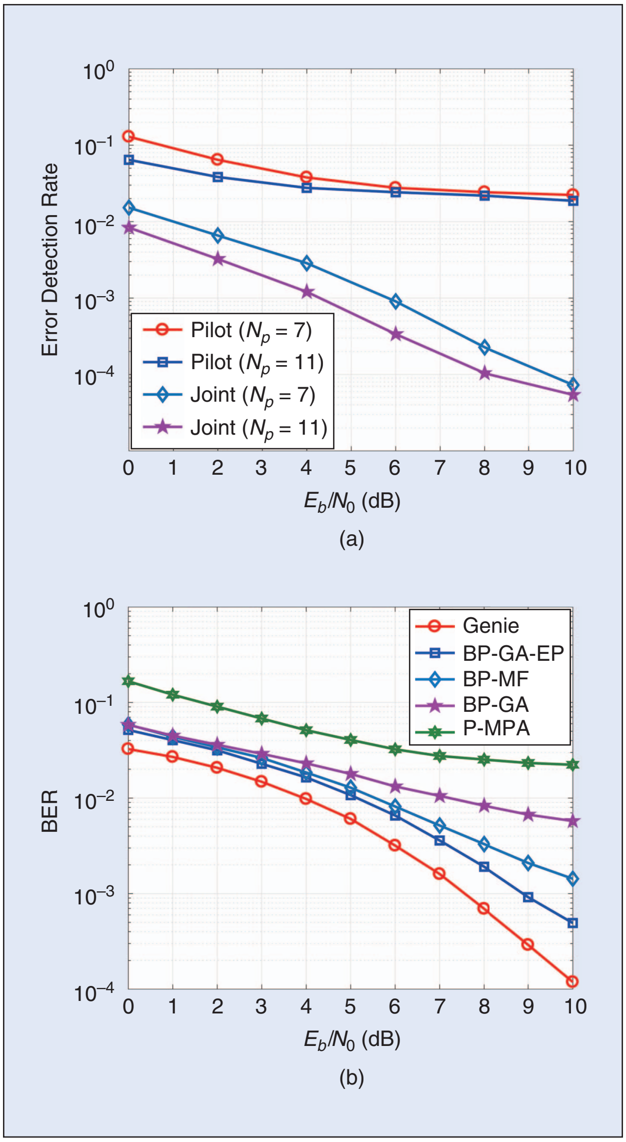


FIGURE 8 An evaluation of (a) the error-detection rate by joint and pilot-only methods and (b) the BER performance of the four-ary SCMA system. BP-GA-EP denotes the hybrid BP and EP message passing with Gaussian approximation. BP-MF represents the hybrid BP and mean-field approximation. BP-GA is the BP Gaussian approximation using the central limit theorem.

Conclusions and Future Challenges

This article reviewed the state of the art for SCMA technologies. We specifically considered techniques using a code-domain NOMA, SCMA, as a promising for wireless interfaces of the future. We showed how a proper mother constellation design as well as the dynamic FG construction can lead to an improved SCMA codebook design. This article also described issues related to low-complexity decoders accommodating multiuser decoding for SCMA. The idea is based on sphere decoding, which avoids exhaustive searching during the MPA. To meet the URLLC requirement, SCMA systems adopt grant-free transmission methods. We introduced a receiver that performs joint channel estimation, data decoding, and the detection of active users based on EP message passing. Finally, a UE-to-CTU mapping rule based on ACK feedback, a solution to the user-collision problem.

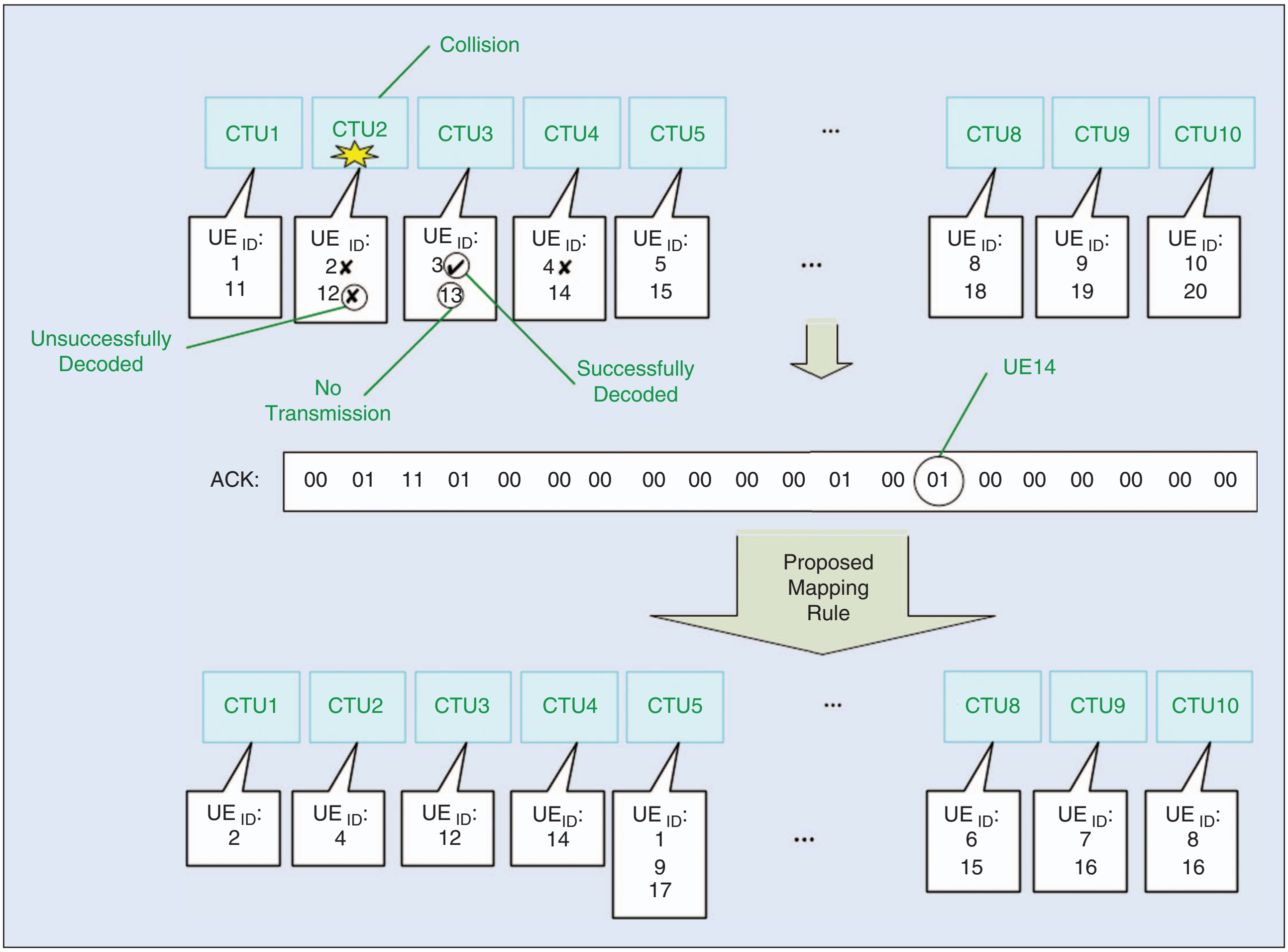


FIGURE 9 The mapping rule based on ACK feedback.

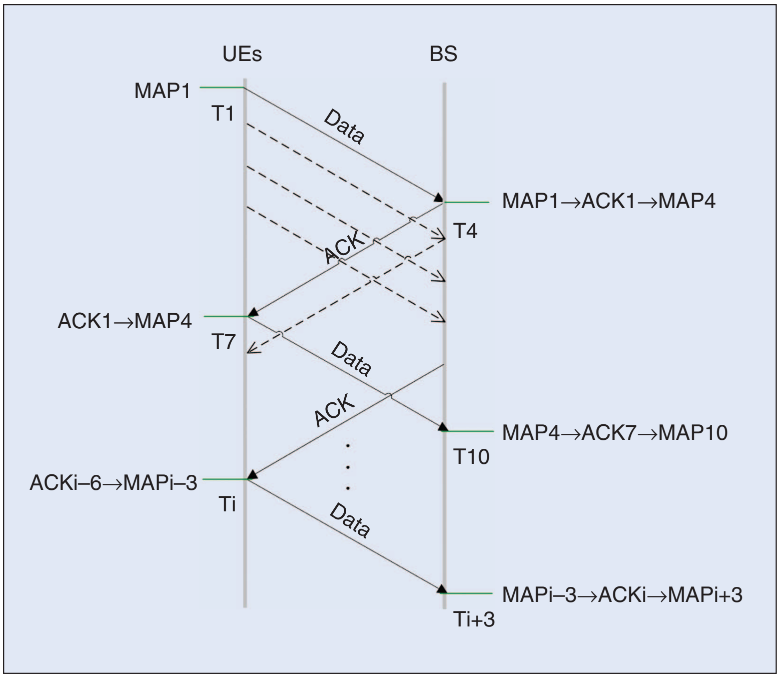


FIGURE 10 The procedure of the grant-free transmission with the ACK feedback-based method.

Issues related to SCMA that require further investigation include the following:

- The decoding of SCMA relies on the loopy BP algorithm. However, conventional BP exhibits poor convergence behavior on the loopy FGs. The convergence rate is slow, and algorithms occasionally achieved local optimal points. BP rules are the consequence of minimizing the constrained Bethe free energy, which is no longer convex when FG contains loops. To overcome this poor convergence, convex Bethe free energy needs to be constructed for loopy FG. Furthermore, when combined with other 5G techniques, such as massive MIMO, the design of the SCMA decoder should also be recast. The decoding complexity would typically grow exponentially with the number of antennas. Thus, there is an emerging need to develop a low-complexity receiver for massive MIMO SCMA.
- Since multiple users share the same resource element in NOMA, the user pairing

serves as another interest problem. For power-domain NOMA, with fixed power allocation, the sum rate can be enlarged by selecting users with distinctive channel conditions. On the other hand, it is better to pair the users with similar channel conditions in cognitive radio-inspired NOMA [15]. Unlike power-domain NOMA that transmits on a single resource element, the situation for SCMA is rather complex when the user paring is performed on multiple carriers.

- Artificial intelligence (AI) combined with 5G is another research trend. AI can be used to solve problems that are intractable in a classical communication system. For instance, combinatorial optimization is typically the NP-hard problem and is frequently encountered in SCMA, such as in issues related to codebook assignment, user selection, and resource allocation. Alternative solutions to those problems might be found by AI using machine-learning-based methods. The SCMA system can also be formulated as an autoencoder, where the end-to-end communication can be joint-optimized through the construction of deep neural networks. The autoencoder is trained through the BER or block error rate and can accommodate various channel conditions as well as system parameters. The performance of SCMA would likely improve with the much lower computational complexity.

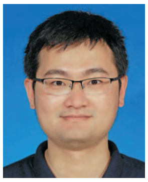
Acknowledgments

This work was supported by the National Natural Science Foundation of China under grants 61671294 and 61701301, the Science and Technology Commission of Shanghai Municipality Project under grants 16JC1402900 and 17510740700, the Natural Science Foundation of Guangxi Province under grant 2015GXNSFDA139037, and the National Science and Technology Major Project under grants 2017ZX03001002-005 and 2018ZX03001009-002.

Author Information

Fan Wei (weifan89@sjtu.edu.cn) is currently pursuing his Ph.D. degree in information and communication engineering at the Shanghai Institute for Advanced Communication and Data Science, Department of Electronic Engineering, Shanghai Jiao Tong University, China. His research interests include wireless communication with a focus on nonorthogonal multiple access.

Wen Chen (wenchen@sjtu.edu.cn) is a professor at the Shanghai Institute for Advanced Communication and Data Science, Department of Electronic Engineering, Shanghai Jiao Tong University, China. His interests include multiple access, coded cooperation, and green heterogeneous networks.



Yongpeng Wu (yongpeng.wu@sjtu.edu.cn) is an associate professor at the Shanghai Institute for Advanced Communication and Data Science, Department of Electronic Engineering, Shanghai Jiao Tong University, China. His research interests include massive multiple-input, multiple-output (MIMO) systems; massive MIMO systems; and physical-layer security.

Jun Li (jun.li@njust.edu.cn) is a professor at the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, China. His research interests include network information theory, ultradense wireless networks, and mobile edge computing.

Yuan Luo (yuanluo@sjtu.edu.cn) is a professor in the Computer Science and Engineering Department, Shanghai Jiao Tong University, China. His research interests include information theory, communication coding theory, and computer security.

References

[1] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE 24th Int. Symp. Personal Indoor and Mobile Radio Communications (PIMRC)*, pp. 332–336, 2013.

[2] L. Yu, X. Lei, P. Fan, and D. Chen, "An optimized design of SCMA codebook based on star-QAM signaling constellations," in *Proc. Int. Conf. Wireless Communications and Signal Processing (WCSP)*, pp. 1–5, Oct. 2015.

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

[4] Z. Li, W. Chen, F. Wei, F. Wang, X. Xu, and Y. Chen, "Joint codebook assignment and power allocation for SCMA based on capacity with Gaussian input," in *Proc. IEEE Int. Conf. Communications in China (ICCC)*, Chengdu, China, July 2016, pp. 1–6.

[5] D. Cai, P. Fan, and P. T. Mathiopoulos, "A tight lower bound for the symbol error performance of the uplink sparse code multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 190–193, Apr. 2017.

[6] J. Bao, Z. Ma, M. Xiao, Z. Ding, and Z. Zhu, "Performance analysis of uplink SCMA with receiver diversity and randomly deployed users," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2792–2797, Mar. 2016.

[7] Y. Wu, S. Zhang, and Y. Chen, "Iterative multiuser receiver in sparse code multiple access systems," in *Proc. IEEE Int. Conf. Communications (ICC)*, London, 2015, pp. 2918–2923.

[8] H. Mu, Z. Ma, M. Alhaji, P. Fan, and D. Chen, "A fixed low complexity message pass algorithm detector for up-link SCMA system," *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 585–588, Dec. 2015.

[9] A. Bayesteh, H. Nikopour, M. Taherzadeh, H. Baligh, and J. Ma, "Low complexity techniques for SCMA detection," in *Proc. IEEE Globecom Workshops*, Dec. 2015, pp. 1–6.

[10] F. Wei and W. Chen, "A low complexity SCMA decoder based on list sphere decoding," in *Proc. IEEE Global Commun.*, Washington, D.C., Dec. 2016.

[11] F. Wei and W. Chen, "Low complexity iterative receiver design for sparse code multiple access," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 621–634, Feb. 2017.

[12] A. Bayesteh, E. Yi, H. Nikopour, and H. Baligh, "Blind detection of SCMA for uplink grant-free multiple-access," in *Proc. Int. Symp. Wireless Wireless Communication Systems*, Aug. 2014, pp. 853–857.

[13] F. Wei, W. Chen, Y. Wu, J. Ma, and T. A. Tsiftsis, "Message-passing receiver design for joint channel estimation and data decoding in uplink grant-free SCMA systems," *IEEE Trans. Wireless Commun.*, to be published.

[14] J. Shen, W. Chen, F. Wei, and Y. Wu, "ACK feedback based UE-to-CTU mapping rule for SCMA uplink grant-free transmission," in *Proc. Int. Conf. Wireless Communications and Signal Processing (WCSP)*, Nanjing, China, Oct. 2017, pp. 1–5.

[15] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

VT