Non-Orthogonal Multiple Access for 5G: Solutions, Challenges, Opportunities, and Future Research Trends

Linglong Dai, Bichai Wang, Yifei Yuan, Shuangfeng Han, Chih-Lin I, and Zhaocheng Wang

ABSTRACT

The increasing demand of mobile Internet and the Internet of Things poses challenging requirements for 5G wireless communications, such as high spectral efficiency and massive connectivity. In this article, a promising technology, non-orthogonal multiple access (NOMA), is discussed, which can address some of these challenges for 5G. Different from conventional orthogonal multiple access technologies, NOMA can accommodate much more users via non-orthogonal resource allocation. We divide existing dominant NOMA schemes into two categories: power-domain multiplexing and code-domain multiplexing, and the corresponding schemes include power-domain NOMA, multiple access with low-density spreading, sparse code multiple access, multi-user shared access, pattern division multiple access, and so on. We discuss their principles, key features, and pros/cons, and then provide a comprehensive comparison of these solutions from the perspective of spectral efficiency, system performance, receiver complexity, and so on. In addition, challenges, opportunities, and future research trends for NOMA design are highlighted to provide some insight on the potential future work for researchers in this field. Finally, to leverage different multiple access schemes including both conventional OMA and new NOMA, we propose the concept of software defined multiple access (SoDeMA), which enables adaptive configuration of available multiple access schemes to support diverse services and applications in future 5G networks.

INTRODUCTION

In the history of wireless communications from the first generation (1G) to 4G, the multiple access scheme has been the key technology to distinguish different wireless systems. It is well known that frequency-division multiple access (FDMA) for 1G, time-division multiple access (TDMA) mostly for 2G, code-division multiple access (CDMA) for 3G, and orthogonal frequency-division multiple access (OFDMA) for 4G are primarily orthogonal multiple access (OMA) schemes. In these conventional multiple access schemes, different users are allocated to orthogonal resources in either the time, frequency, or code domain in order to avoid or alleviate inter-user interference. In this way, multiplexing gain can be achieved with reasonable complexity.

However, the fast growth of mobile Internet has propelled 1000-fold data traffic increase by 2020 for 5G. Hence, the spectral efficiency becomes one of the key challenges to handle such explosive data traffic. Moreover, due to the rapid development of the Internet of Things (IoT), 5G needs to support massive connectivity of users and/or devices to meet the demand for low latency, low-cost devices, and diverse service types. To satisfy these requirements, enhanced technologies are necessary. So far, some potential candidates have been proposed to address challenges of 5G, such as massive MIMO, millimeter wave communications, ultra dense network, and non-orthogonal multiple access (NOMA) [1]. In this article, we focus on NOMA, which is highly expected to increase system throughput and accommodate massive connectivity. Note that Third Generation Partnership Project (3GPP) Long Term Evolution (LTE) Rel-13 is doing ongoing studies toward NOMA in the form of multi-user superposition transmission (MUST). NOMA allows multiple users to share time and frequency resources in the same spatial layer via power domain or code domain multiplexing. Recently, several NOMA schemes have attracted lots of attention, and we can generally divide them into two categories,[1] that is, power domain multiplexing [2–4] and code domain multiplexing, including multiple access with low-density spreading (LDS) [5, 6], sparse code multiple access (SCMA) [7], multi-user shared access (MUSA) [8], and so on. Some other multiple access schemes such as pattern-division multiple access (PDMA) and bit division multiplexing (BDM) [9] are also proposed. Key features and advantages of NOMA are discussed

Linglong Dai, Bichai Wang, and Zhaocheng Wang are with Tsinghua University.

Yifei Yuan is with ZTE Corporation.

Shuangfeng Han and Chih-Lin I are with China Mobile Research Institute.

This work was supported in part by the International Science & Technology Cooperation Program of China (Grant No. 2015DFG12760), the National Natural Science Foundation of China (Grant Nos. 61571270 and 61201185), and the Beijing Natural Science Foundation (Grant No. 4142027)..

[1] Note that "NOMA" is also used by NTT DoCoMo to refer to NOMA via power domain multiplexing.

0163-6804/15/$25.00 © 2015 IEEE

later. The design principles, key features, advantages and disadvantages of existing dominant NOMA schemes are discussed and compared. More importantly, although NOMA can provide attractive advantages, some challenging problems should be solved, such as advanced transmitter design and the trade-off between performance and receiver complexity. Thus, opportunities and research trends are highlighted to provide some insights on the potential future work for researchers in this field. In addition, unlike the conventional way of designing a specific multiple access scheme separately and individually, we propose the concept of software defined multiple access (SoDeMA), in which several candidates among multiple access schemes can be adaptively configured to satisfy different requirements of diverse services and applications in future 5G networks. Finally, conclusions are drawn.

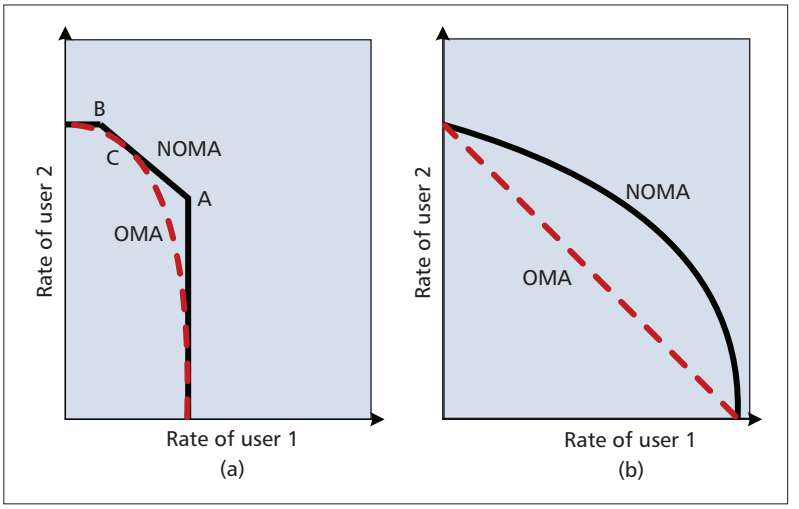

Figure 1. Channel capacity comparison of OMA and NOMA in an AWGN channel: a) uplink AWGN channel; b) downlink AWGN channel.

Features of NOMA

In conventional OMA schemes, multiple users are allocated with radio resources which are orthogonal in time, frequency, or code domain. Ideally, no interference exists among multiple users due to the orthogonal resource allocation in OMA, so simple single-user detection can be used to separate different users' signals. Theoretically, it is known that OMA cannot always achieve the sum-rate capacity of multiuser wireless systems [10]. Apart from that, in conventional OMA schemes, the maximum number of supported users is limited by the total amount and the scheduling granularity of orthogonal resources.

Recently, NOMA has been investigated to deal with the problems of OMA as mentioned above. Basically, NOMA allows controllable interferences by non-orthogonal resource allocation with the tolerable increase in receiver complexity. Compared to OMA, the main advantages of NOMA include the following.

Improved spectral efficiency: According to the multi-user capacity analysis in the pioneering work [10], Fig. 1 shows the channel capacity comparison of OMA and NOMA, where two users in the additive white Gaussian noise (AWGN) channel are considered as an example without loss of generality. Figure 1a shows that the uplink NOMA is able to achieve the capacity bound, while OMA schemes are in general suboptimal except at point C. However, at this optimal point, the user throughput fairness is quite poor when the difference of the received powers of the two users is significant, as the rate of the weak user is much lower than that of the strong user. In the downlink, Fig. 1b shows that the boundary of rate pairs of NOMA is outside of the OMA rate region in general. In multi-path fading channels with intersymbol interference (ISI), although OMA could achieve the sum capacity in the downlink, NOMA is optimal while OMA is strictly suboptimal if channel state information (CSI) is only known at the mobile receiver [10].

Massive connectivity: The non-orthogonal resource allocation in NOMA indicates that the number of supported users or devices is not strictly limited by the amount of available resources and their scheduling granularity. Therefore, NOMA can accommodate significantly more users than OMA by using non-orthogonal resource allocation; for example, MUSA can still achieve reasonably good performance when the overloading is 300 percent [8].

Low transmission latency and signaling cost: In conventional OMA with grant-based transmission, a user has to send a scheduling request to the base station (BS) at first. Then, based on the received request, the BS performs scheduling for the uplink transmission and sends a grant over the downlink channel. This procedure results in large latency and high signaling cost, which becomes worse or even unacceptable in the scenario of massive connectivity anticipated for 5G. In contrast, such dynamic scheduling is not required in some uplink schemes of NOMA, rendering a grant-free uplink transmission that can drastically reduce the transmission latency and signaling overhead.

Due to the potential advantages above, NOMA has been actively investigated as a promising technology for 5G. In the next section, existing dominant NOMA schemes are discussed and compared in detail.

Dominant NOMA Solutions

In this section, we discuss dominant NOMA schemes by grouping them into two categories: power domain multiplexing and code domain multiplexing. Power domain multiplexing means that different users are allocated different power levels according to their channel conditions to obtain the maximum gain in system performance. Such power allocation is also beneficial to separate different users, where successive interference cancellation (SIC) is often used to cancel multi-user interference. In this article, power domain multiplexing is applied only to downlink NOMA. Code domain multiplexing is similar to CDMA or multicarrier CDMA (MC-CDMA), that is, different users are assigned different codes, and are then multiplexed over the same time-frequency resources. The difference

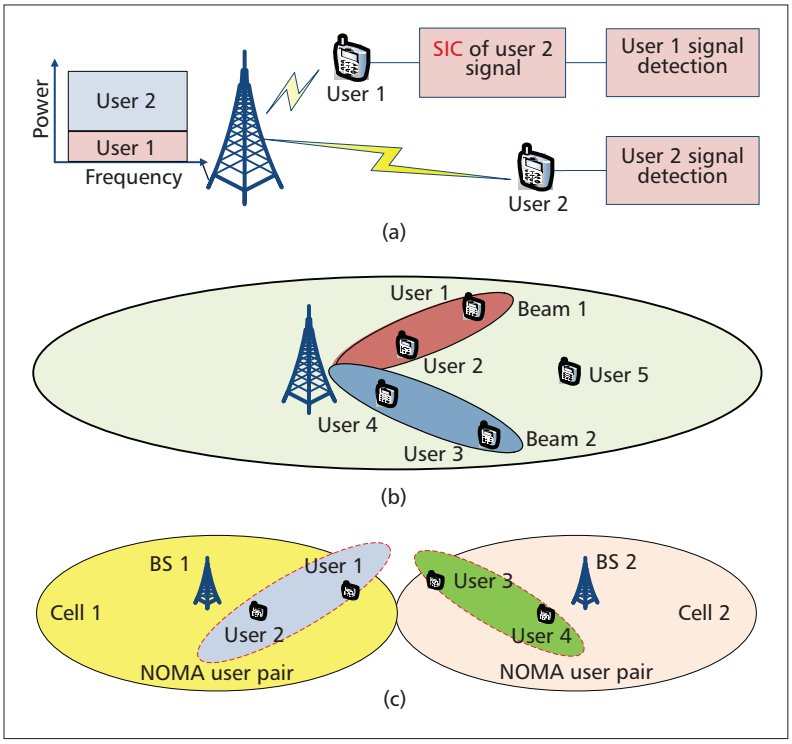

Figure 2. Illustration of NOMA via power domain multiplexing: a) basic NOMA with a SIC receiver; b) NOMA in MIMO systems; c) network NOMA.

between power domain multiplexing and code domain multiplexing is that code domain multiplexing can achieve certain spreading gain and shaping gain at the cost of increased signal bandwidth.

NOMA via Power Domain Multiplexing

Basic NOMA with a SIC Receiver: Figure 2a illustrates the basic NOMA scheme via power domain multiplexing with a SIC receiver in the downlink. Note that this NOMA scheme can also be applied in the uplink [2]. At the BS transmitter, signals for different users are linearly added up under certain power partitions to balance the sum rate of all multiplexed users and the throughput fairness among individual users.

At the receiver, SIC is commonly used to realize multi-user detection (MUD). Due to the near-far effect, the channel conditions may vary significantly among users. SIC is performed at users with relatively high signal-to-interference-plus-noise ratio (SINR), and should be carried out in descending order of SINR.

As we can see, the basic form of NOMA with SIC exploits SINR difference among users, either due to the natural near-far effect or by non-uniform power allocation at the transmitter. A similar scheme can be used for uplink to increase the uplink system capacity.

NOMA in Massive MIMO Systems: NOMA can be used in conjunction with multi-user multiple-input multiple-output (MU-MIMO) to further improve the system spectral efficiency [3]. As illustrated in Fig. 2b, multiple transmit antennas at a BS are used to form different beams in the spatial domain, where each beam adopts the basic NOMA discussed above.

At the receiver, the inter-beam interference can be suppressed by spatial filtering [3], and then intra-beam SIC can be used to remove the inter-user interference. The extension of NOMA in massive MIMO systems can further improve spectral efficiency.

Network NOMA: When transmit power allocation is biased toward far away users in downlink NOMA, cell edge users experience increased interference from neighboring cells. As an example, a cellular system with two cells and four users is depicted in Fig. 2c, where a two-user NOMA scheme is assumed: user 1 and user 2 are served by BS 1, while user 3 and user 4 are served by BS 2. Strong interference is expected between users 1 and 3, which may degrade the performance of network NOMA, that is, multi-cell NOMA.

To mitigate the inter-cell interference, joint precoding of NOMA users' signals across neighboring cells can be considered. This requires that all users' data and CSI should be available at multiple BSs, but finding the optimal precoder is not trivial. Moreover, the multi-user precoding used for single-cell NOMA maybe not be feasible for the network NOMA scenario, since the precoder for geographically separate BS antennas does not actually form the physical beam that can readily be used for intra-beam NOMA. Based on the fact that large-scale fading would be quite different between the links to different cells, a complexity-reduced precoding scheme for network NOMA has been proposed in [4], where the multi-cell joint precoder is applied only to cell edge users (e.g., user 1 and user 3 as shown in Fig. 2c).

NOMA via Code Domain Multiplexing

Low-Density Spreading CDMA: The idea behind LDS-CDMA is to use sparse spreading sequences instead of dense spreading sequences in conventional CDMA [5] to reduce the interference at each chip. Therefore, LDS-CDMA can improve system performance by exploiting LDS sequences in CDMA [5], which is the key feature distinguishing conventional CDMA and LDS-CDMA. In this way, interference will be efficiently decreased among multiple users with appropriate spreading sequence design, and overloading can be achieved.

At the receiver, a message passing algorithm (MPA) can be used to realize MUD. MPA is very efficient for the factor graph [11], which is a bipartite graph including variable nodes and factor nodes as illustrated in Fig. 3. Messages are passed among variable nodes and factor nodes over edges, which can be interpreted as the soft-values that represent the reliability of the symbol associated with each edge. The marginal distribution of a variable node can be regarded as a function of the messages received by that node [11].

Low-Density Spreading OFDM: LDS orthogonal frequency-division multiplexing (LDS-OFDM) can be considered as a combined version of LDS-CDMA and OFDM, in which

the chips are subcarriers of OFDM in order to combat the multipath fading. In LDS-OFDM, the transmitted symbols are first mapped to certain LDS sequences, and then transmitted on different OFDM subcarriers. The number of symbols can be greater than the number of subcarriers, that is, overloading is allowed to improve spectral efficiency [6]. MPA in LDS-CDMA can also be used in an LDS-OFDM receiver. Essentially, LDS-OFDM can be viewed as an improved form of multi-carrier CDMA (MC-CDMA) by replacing the dense spreading sequences with LDS.

Sparse Code Multiple Access: The recently proposed SCMA [7] is an enhanced version of LDS-CDMA. Unlike LDS-CDMA, SCMA directly maps different bitstreams to different sparse codewords, as illustrated in Fig. 4, where each user has a predefined codebook (there are 6 users in Fig. 4). All codewords in the same codebook contain zeros in the same two dimensions, and the positions of zeros in different codebooks are distinct to facilitate the collision avoidance of any two users. For each user, two bits are mapped to a complex codeword. Codewords for all users are multiplexed over four shared orthogonal resources (e.g., OFDM subcarriers).

The key difference between LDS-CDMA and SCMA is that a multi-dimensional constellation for SCMA is designed to generate codebooks, which brings the "shaping gain" that is not possible for LDS [7]. Here, "shaping gain" is the gain in the average symbol energy when the shape of a constellation is changed. In general, the shaping gain is higher when the shape of a constellation is closer to a sphere, and the maximum achievable shaping gain by the optimization of a multi-dimensional constellation is 1.53 dB [7]. For the concatenated approach in high modulation order, the multi-dimensional constellation can be optimized to obtain shaping gain, and then codebooks are generated based on the multi-dimensional constellation [7]. The SCMA codebook design is a complicated problem, since different layers are multiplexed with different codebooks. As the appropriate design criterion and specific solution to the multi-dimensional problem are still unknown, a multi-stage approach has been proposed to realize a suboptimal solution [7]. Specifically, an N-dimensional complex constellation with M points (which is called the mother constellation) is first optimized to improve the shaping gain, and then some codebook-specific operations are performed to the mother constellation to generate the N-dimensional constellation for each codebook. Three typical operations are phase rotation, complex conjugate, and dimensional permutation of the constellation [7]. In the generated N-dimensional constellations after codebook-specific operations, each N-dimensional constellation point is multiplied with a projection matrix to generate a K-dimensional codeword ($K \gg N$), which has N non-zero elements from the components of the N-dimensional constellation point. In this way, codebooks with M codewords can be obtained. Readers can find more details in [7].

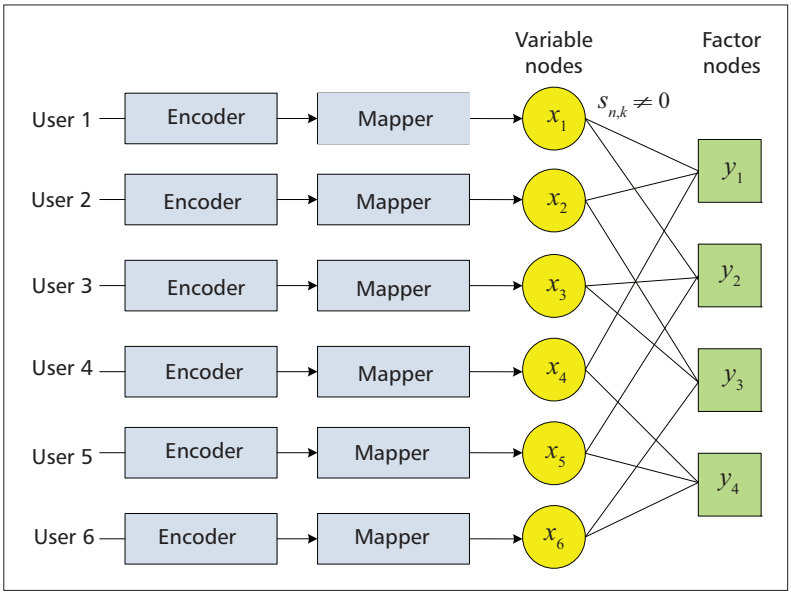

Figure 3. Illustration of LDS-CDMA: six users and four chips for transmission, which means 150 percent overloading.

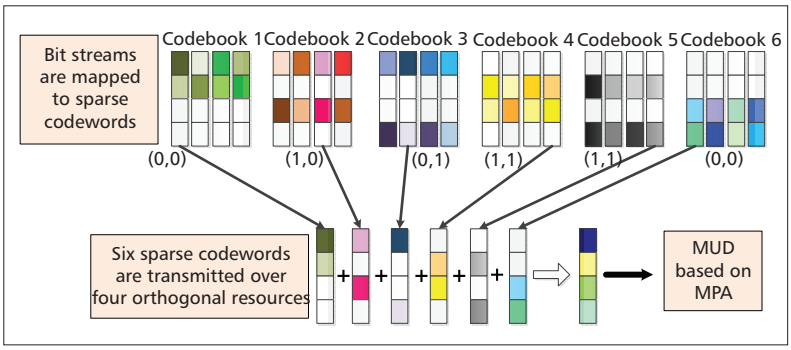

Figure 4. SCMA encoding and multiplexing.

Multi-User Shared Access: In the uplink MUSA system shown in Fig. 5 [8], symbols of each user are spread by a spreading sequence. Multiple spreading sequences constitute a pool from which each user can randomly pick one of the sequences. Note that for the same user, different spreading sequences may also be used for different symbols, which may further improve the performance via interference averaging. Then all spreading symbols are transmitted over the same time-frequency resources. The spreading sequences should have low cross-correlation and can be M-ary. At the receiver, codeword-level SIC is used to separate data from different users. The complexity of codeword-level SIC is less of an issue in the uplink as in any case the receiver needs to decode the data for all users. The only noticeable impact on the receiver implementation would be that the pipeline of processing may be changed in order to perform SIC operation. The difference between MUSA and MC-CDMA is that MUSA assumes that it is basically synchronous when users' signals arrive at the BS, which is easier to realize SIC, while MC-CDMA dose not require this synchronization in the uplink. In addition, MUSA uses non-binary spreading sequences, while binary

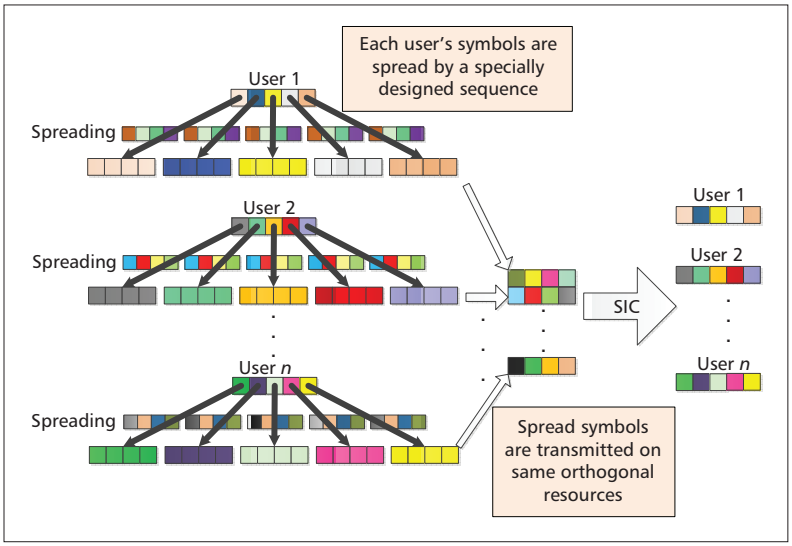

Figure 5. Uplink MUSA system.

spreading sequences are usually considered in classical MC-CDMA systems.

In downlink MUSA, users are separated into K groups. In each group, different users' symbols are mapped to different constellations in a way that can ensure Gray mapping in the combined constellation of superposed signals. The combined constellation is determined not only by the modulation order of each user, but also by the transmit power partition among multiplexed users. Orthogonal sequences can be used to spread the superposed symbols to get time or frequency diversity gain. Gray mapping of the combined constellation reduces the reliance on advanced receivers, so less processing-intensive receivers such as symbol-level SIC can be used.

Other NOMA Schemes

In addition to the power domain multiplexing and code domain multiplexing discussed above, a few other NOMA schemes are currently being investigated. Pattern-division multiple access (PDMA) is a NOMA scheme that can be realized in several domains. At the transmitter, PDMA uses non-orthogonal patterns, which are designed by maximizing the diversity and minimizing the overlaps among multiple users. Then the actual multiplexing can be carried out in the code domain, spatial domain, or a combination of the two. Multiplexing in the code domain corresponds to the case of successive interference cancellation amenable multiple access (SAMA) [12], which is similar to LDS-CDMA, with LDS sequences being replaced by non-orthogonal patterns. Hence, MPA can also be used for the sequence detection in PDMA. The multiplexing in the spatial domain, called spatial PDMA, requires multiple antennas at the BS. The diversity of PDMA can come from multiple transmit antennas, which is preferred for macrocell deployment. Different from multi-user MIMO, precoding is not needed in spatial PDMA since the aim is to increase the spatial diversity rather than spectral efficiency. PDMA can be used in both downlink and uplink transmissions.

Bit-division multiplexing (BDM) [9] is another form of NOMA particularly useful for downlink transmission. Its basic concept is based on hierarchical modulation, and the resources of multiplexed users are partitioned at the bit level. Although strictly speaking the resource allocation of BDM is orthogonal in the bit domain, multi-user signals share the same constellation (e.g., superposed in the modulation symbol domain).

Some other NOMA schemes were also proposed, such as interleave-division multiple access (IDMA), which performs interleaving of chips after symbols are multiplied by spreading sequences. As shown in [13], compared to CDMA, IDMA is able to achieve an E_b/N_0 gain of about 1 dB when bit error rate (BER) performance of 10^{-3} is considered in highly loaded systems with 200 percent overloading.

In many of the NOMA schemes mentioned above, especially when used for grant-free uplink transmission, there is an issue that the users' activity or instantaneous system loading is not readily known to the receiver. This would have a negative impact on the performance. Compressive sensing (CS) is a promising technique to estimate the resource occupancy. Some work on CS-based random access has been carried out recently, such as compressive random access [14].

Comparison of NOMA Solutions

From the theoretical perspective, code-domain NOMA can obtain spreading gain due to the use of spreading sequences or codewords, which can be achieved only in the case that there is no CSI at the transmitter. Spreading gain is similar to that in CDMA, that is, the transmitted bandwidth can be spread by spreading sequences or codewords, and thus, according to Shannon's equation, signals can still be transmitted with the same capacity even when signal-to-noise ratio (SNR) is low. The spreading gain can be calculated by 10log N, where N is the spreading factor. However, introducing redundancy through spreading will affect the system spectral efficiency [15]. In addition, SCMA can achieve extra "shaping gain" due to the optimization of multi-dimensional constellation [7].

We also compare these NOMA schemes in terms of the computational complexity of the multi-user signal detection algorithm. In power-domain NOMA, SIC is the key method for multi-user interference cancellation with complexity $\mathcal{O}(K^3)$, where K is the number of users. Therefore, the complexity of SIC is much less than that of the optimal maximum likelihood (ML) detection, whose complexity $\mathcal{O}(|\mathbb{X}|^K)$ increases exponentially with the number of users K, where $|\mathbb{X}|$ denotes the cardinality of the constellation set $\mathbb{X}$. On the other hand, in code-domain NOMA like LDS-CDMA, LDS-OFDM, and SCMA, spreading sequences or codebooks should be known at the receiver to realize MUD, and the complexity of the MPA-based receiver is $\mathcal{O}(|\mathbb{X}|^w)$, where w is the maximum number of nonzero signals superimposed on each chip or subcarrier. Thus, an MPA-based receiver usually has higher complexity than a SIC-based receiver as w is usually larger than 3 in typical 5G systems with massive connectivity.

Challenges, Opportunities, and Future Research Trends

Theoretical Analysis of Achievable Rate and Overloading Bounds

In NOMA schemes, theoretical analysis is required to provide some insights for system design. Achievable rate of multiple access is a key metric of system performance. The achievable rate of code-domain NOMA with LDS needs to be studied, and can refer to the analytical approach of MC-CDMA. Particularly, due to the special structure of spreading sequences, some approximations can be used to simplify the calculation. It is expected to derive the closed-form expression to reveal the relationship between the achievable rate and LDS parameters such as sequence sparsity and overloading factor. Such theoretical results can shed light on how to design the system parameters according to the specific application requirements.

On the other hand, the interference cancellation capability and the affordable complexity at the receiver play an important role in the overall performance, for example, the maximum overloading factor that the system can support.

Design of Spreading Sequences or Codebooks

In LDS systems, due to non-orthogonal resource allocation, interference exists among multiple users. A factor graph in MPA should be optimized to get good trade-off between overloading factor and receiver complexity.

In addition, it has been proved that MPA can obtain the exact marginal distribution with a cycle-free factor graph and the precise solution with a local tree-like factor graph. Graph theory can be used to design a cycle-free or local tree-like factor graph in NOMA without compromising spectral efficiency. In addition, the matrix design principle and methods in low-density-parity check (LDPC) can be considered when designing the factor graph for NOMA.

Receiver Design

For an MPA-based receiver, the complexity may still be high for massive connectivity in 5G. Therefore, simplified improvement of MPA can be used to reduce receiver complexity, such as Gaussian approximation of interference (GAI), which models the interference-plus-noise as Gaussian distributed, and such approximation tends to be more accurate as the amount of connectivity becomes larger in 5G. In addition, MPA can be used to jointly detect and decode the received symbols, in which the constructed graph consists of variable nodes, observation nodes, and check nodes corresponding to the check equations of the LDPC code. In this way, intrinsic information between the decoder and the demodulator can be used more efficiently to improve the detector's performance.

For a SIC-based receiver, error propagation may degrade the performance of some users. Therefore, at each stage of SIC, some nonlinear detection algorithms with higher detection accuracy can be considered to suppress the error propagation.

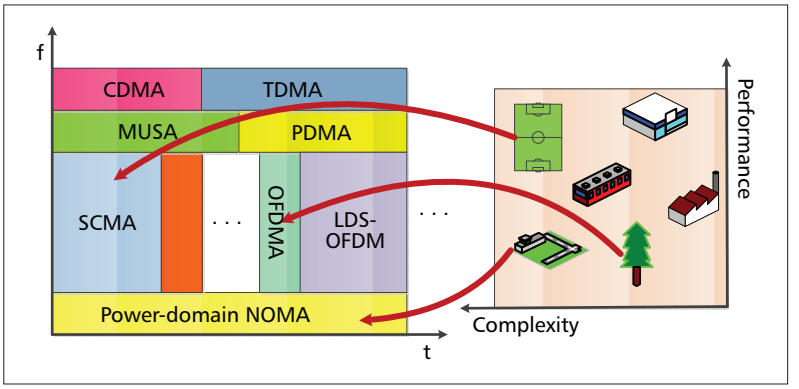

Figure 6. Illustration of the concept of software defined multiple access.

Other Challenges

There are also some other engineering aspects of NOMA, including reference signal design, channel estimation, and CSI feedback mechanism that can deliver robust performance when cross-user interference is severe, resource allocation signaling that can support different transmission modes for NOMA, extension to MIMO (especially massive MIMO) that can reap the performance benefits of both NOMA and multi-user MIMO, peak-to-average-power ratio (PAPR) reduction in multi-carrier NOMA, system scalability that can support different traffic loading and radio environment, and so on. These challenges need to be addressed before NOMA becomes part of 5G standards in the future.

The Concept of SoDeMA

As discussed above, NOMA can be used for capacity improvement and massive connections in 5G. However, this does not mean that conventional OMA schemes will be completely replaced by NOMA in future 5G networks. For example, when the number of users is small and the near-far effect is not significant, such as in the case of small cells, OMA would be a better choice. In this sense, both OMA and NOMA will coexist in 5G to fulfill diverse requirements of different services and applications.

To this end, we borrow the idea of software defined radio (SDR) for multiple access design to propose the SoDeMA concept for 5G as shown in Fig. 6, where different NOMA schemes can coexist in a system assuming all of them will be specified in 5G standards. SoDeMA provides a very flexible configuration of multiple access schemes to support different services and applications in 5G. For example, for cell-center users or real-time services like ultra-high-definition video, conventional OMA schemes can be adopted to support high data rate transmission, which capitalizes on the orthogonality and synchronization. On the other hand, when high spectral efficiency, massive connectivity, and frequent access of small packets are required in some practical scenarios (e.g., dense population areas and mobile social applications), NOMA schemes can be selected. Moreover, different NOMA or OMA schemes have their own appropriate application situations, and can be adaptively configured to realize the trade-off between

Compared with conventional OMA, NOMA allows controllable interferences to realize overloading at the cost of tolerate increase of receiver complexity. Therefore, the demands of spectral efficiency and massive connectivity for 5G can be partially fulfilled by NOMA.

performance and implementation complexity. For instance, if a large difference among users' channel conditions exists due to the near-far effect or in moving networks, power-domain NOMA with a SIC receiver can be used with relatively low complexity. On the other hand, if high reliability should be guaranteed, especially when channel condition is bad or the location distribution of users is concentrated, SCMA is a feasible solution due to its shaping gain and near-optimal MPA detection. Of cause, when the number of users is large enough, it may be difficult to design a codebook for each user, and in this case, LDS-OFDM or MUSA can also be used to reduce the design complexity at the transmitter or receiver, separately.

As elaborated in previous sections, certain signal processing modules are common to several NOMA schemes, for example, MPA at the receiver or spreading operation at the transmitter, which can be shared in hardware so that the hardware cost would be reduced at both user terminals and base stations. These general-purpose modules can be combined in different forms at the software level to implement different schemes. The switching between NOMA schemes is fast and flexible with software defined hardware architecture, and can quickly adapt to different deployment scenarios, that is, from capacity achieving to user loading improvement.

To enable SoDeMA, the frame structure should be flexible enough so that the time and frequency resources are partitioned into different blocks freely for different services and users. In each resource block, one specific multiple access scheme is configured with specific waveform, duplex mode, pilot signals, power level, and so on. Note that the inter-subcarrier interference between different resource blocks needs to be carefully mitigated. The proposed SoDeMA concept provides a flexible configuration of multiple access schemes to support different services and applications. It is highly expected that SoDeMA can be carefully designed to adapt to various application scenarios to support the system design goal of "anything as a service" in future 5G networks.

Conclusions

In this article, we have discussed and compared several major NOMA schemes for 5G from the aspects of basic principles, key features, receiver complexity, engineering feasibility, and so on. Compared to conventional OMA, NOMA allows controllable interferences to realize overloading at the cost of a tolerable increase of receiver complexity. Therefore, the demands of spectral efficiency and massive connectivity for 5G can be partially fulfilled by NOMA. We have also highlighted key challenges, opportunities and future research tends for the design of NOMA, including theoretical work, optimal design of spreading sequences or codebooks, receiver design, a grant-free NOMA mechanism, and so on. The proposed concept of SoDeMA is able to flexibly support diverse services and applications with different requirements. It is expected that NOMA will play an important role in future 5G wireless communications.

References

[1] F. Boccardi *et al.*, "Five Disruptive Technology Directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 74–80.

[2] Y. Saito *et al.*, "Non-Orthogonal Multiple Access (NOMA) for Future Radio Access," *Proc. IEEE VTC-Spring '13*, June 2013, pp. 1–5.

[3] K. Higuchi and Y. Kishiyama, "Non-Orthogonal Access with Random Beamforming and Intra-Beam SIC for Cellular MIMO Downlink," *Proc. IEEE VTC-Fall '13*, Sept. 2013, pp. 1–5.

[4] S. Han *et al.*, "Energy Efficiency and Spectrum Efficiency Co-Design: From NOMA to Network NOMA," *IEEE MMTC E-Letter*, vol. 9, no. 5, Sept. 2014, pp. 21–24.

[5] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel Low-Density Signature for Synchronous CDMA Systems over AWGN Channel," *IEEE Trans. Signal Proc.*, vol. 56, no. 4, Apr. 2008, pp. 1616–26.

[6] M. Al-Imari *et al.*, "Uplink Nonorthogonal Multiple Access for 5G Wireless Networks," *Proc. 11th Int'l. Symp. Wireless Commun. Sys.*, Aug. 2014, pp. 781–85.

[7] H. Nikopour and H. Baligh, "Sparse Code Multiple Access," *Proc. IEEE PIMRC 2013*, Sept. 2013, pp. 332–36.

[8] Z. Yuan, G. Yu, and W. Li, "Multi-User Shared Access for 5G," *Telecommun. Network Technology*, vol. 5, no. 5, May 2015, pp. 28–30.

[9] J. Huang *et al.*, "Scalable Video Broadcasting Using Bit Division Multiplexing," *IEEE Trans. Broadcast.*, vol. 60, no. 4, Dec. 2014, pp. 701–06.

[10] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge Univ. Press, 2005.

[11] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor Graphs and the Sum-Product Algorithm," *IEEE Trans. Info. Theory*, vol. 47, no. 2, Feb. 2001 , pp. 498–519.

[12] X. Dai *et al.*, "Successive Interference Cancelation Amenable Multiple Access (SAMA) for Future Wireless Communications," *Proc. IEEE ICCS 2014*, Nov. 2014, pp. 1–5.

[13] K. Kusume, G. Bauch, and W. Utschick, "IDMA vs. CDMA: Analysis and Comparison of Two Multiple Access Schemes," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 78–87, Jan. 2012.

[14] G. Wunder, P. Jung, and C. Wang, "Compressive Random Access for Post-LTE Systems," *Proc. IEEE ICC '14*, June 2014, pp. 539–44.

[15] V. V. Veeravalli and A. Mantravadi, "The Coding-Spreading Tradeoff in CDMA Systems," *IEEE JSAC*, vol. 20, no. 2, Feb. 2002, pp. 396–408.

Biographies

Linglong Dai [M'11, SM'14] (daill@tsinghua.edu.cn) received his B.S. degree from Zhejiang University in 2003, his M.S. degree (with highest honor) from the China Academy of Telecommunications Technology (CATT) in 2006, and his Ph.D. degree (with the highest honor) from Tsinghua University, Beijing, China, in 2011. From 2011 to 2013, he was a postdoctoral fellow with the Department of Electronic Engineering, Tsinghua University, where he has been an assistant professor since July 2013. His research interests are in wireless communications, with a focus on multi-carrier techniques, multi-antenna techniques, and multi-user techniques. He has published over 60 journal and conference papers. He has received the Outstanding Ph.D. Graduate of Tsinghua University award in 2011, the Excellent Doctoral Dissertation of Beijing award in 2012, the IEEE ICC Best Paper Award in 2013, the National Excellent Doctoral Dissertation Nomination Award in 2013, the IEEE ICC Best Paper Award in 2014, the URSI Young Scientists Award in 2014, and the IEEE Scott Helt Memorial Award in 2015 (*IEEE Transactions on Broadcasting* Best Paper Award). He currently serves as Co-Chair of the IEEE Special Interest Group (SIG) on Signal Processing Techniques in 5G Communication Systems.

Bichai Wang [S'15] (wang-bc11@mails.tsinghua.edu.cn) received her B.S. degree in electronic engineering from Tsinghua University in 2015. She is currently working toward her Ph.D. degree in the Department of Electronic Engineering, Tsinghua University. Her research interests are in wireless communications, with emphasis on new multiple access techniques. She received the Freshman Scholarship of Tsinghua University in 2011, Academic Merit Scholarships of Tsinghua University in 2012, 2013, and 2014, respectively, and the Excellent Thesis Award of Tsinghua University in 2015.

YIFEI YUAN (yifei.yuan@ztetx.com) received Bachelor's and Master's degrees from Tsinghua University, and a Ph.D. from Carnegie Mellon University, Pennsylvania He was with Alcatel-Lucent from 2000 to 2008 working on 3G/4G key technologies. Since 2008, he has been with ZTE, responsible for standards research on LTE-Advanced physical layer and 5G technologies. His research interests include MIMO, iterative codes, resource scheduling, non-orthogonal access, and small cells. He was admitted to the Thousand Talent Plan Program of China in 2010. He has published extensively, including a book on LTE-A relay and a book on LTE-Advanced key technologies. He has over 30 granted patents.

SHUANGFENG HAN (hanshuangfeng@chinamobile.com) received his M.S. and Ph.D. degrees in electrical engineering from Tsinghua University in 2002 and 2006, respectively. He joined Samsung Electronics as a senior engineer in 2006 working on MIMO, multi-BS MIMO, and so on. Since 2012, he has been a senior project manager in the Green Communication Research Center at the China Mobile Research Institute. His research interests are green 5G, massive MIMO, full duplex, NOMA, and EE-SE co-design.

CHIH-LIN I (icl@chinamobile.com) received her Ph.D. degree in electrical engineering from Stanford University. She has been working at multiple world-class companies and research institutes leading R&D, including AT&T Bell Labs, AT&T HQ, ITRI of Taiwan, and ASTRI of Hong Kong. She received the *IEEE Transactions on Communications* Stephen Rice Best Paper Award and is a winner of the CCCP National 1000 Talent program. Currently, she is China Mobile's chief scientist of wireless technologies and has established the Green Communications Research Center, spearheading major initiatives including system architectures, technologies, and devices; green energy; and C-RAN and soft base stations. She was an elected Board Member of IEEE ComSoc, Chair of the ComSoc Meetings and Conferences Board, and Founding Chair of the IEEE WCNC Steering Committee. She is currently an Executive Board Member of GreenTouch and a Network Operator Council Member of ETSI NFV. Her research interests are green communications, C-RAN, network convergence, and bandwidth active antenna arrays.

ZHAOCHENG WANG [M'09, SM'11] (zcwang@tsinghua.edu.cn) received his B.S., M.S., and Ph.D. degrees from Tsinghua University in 1991, 1993, and 1996, respectively. From 1996 to 1997, he was a postdoctoral fellow of Nanyang Technological University, Singapore. From 1997 to 1999, he was with OKI Techno Centre (Singapore) Pte. Ltd., where he was first a research engineer and later became a senior engineer. From 1999 to 2009, he was with Sony Deutschland GmbH, where he was first a senior engineer and later became a principal engineer. He is currently a professor with the Department of Electronic Engineering, Tsinghua University, and serves as director of the Broadband Communication Key Laboratory, Tsinghua National Laboratory for Information Science and Technology. He has authored or coauthored over 80 international journal papers (SCI indexed). He is the holder of 34 granted U.S./EU patents. He co-authored two books, one of which, *Millimeter Wave Communication Systems*, was selected for the IEEE Series on Digital & Mobile Communication and published by Wiley-IEEE Press. His research areas include wireless communications, visible light communications, millimeter-wave communications, and digital broadcasting. He is a Fellow of the Institution of Engineering and Technology. Currently he serves as an Associate Editor of *IEEE Transaction on Wireless Communications* and *IEEE Communications Letters*, and has also served as Technical Program Committee Co-Chair of various international conferences.