SCIENCE CHINA
Information Sciences

• PERSPECTIVE •

June 2024, Vol. 67, Iss. 6, 166301:1–166301:3
https://doi.org/10.1007/s11432-024-4014-x

Next generation multiple access for IMT towards 2030 and beyond

Zhiguo DING[1,2*], Robert SCHOBER[3], Pingzhi FAN[4] & H. Vincent POOR[5]

[1]*Department of Computer and Information Engineering, Khalifa University, Abu Dhabi, UAE;*
[2]*Department of Electrical and Electronic Engineering, University of Manchester, Manchester* M1 9BB*, UK;*
[3]*Institute for Digital Communications, Friedrich Alexander-University Erlangen-Nurnberg (FAU), Erlangen* 91054*, Germany;*
[4]*CSNMT International Cooperation Research Center (MoST), Southwest Jiaotong University, Chengdu* 610032*, China;*
[5]*Department of Electrical and Computer Engineering, Princeton University, Princeton* NJ 08544*, USA*

Received 4 April 2024/Revised 17 April 2024/Accepted 22 April 2024/Published online 22 May 2024

Multiple access techniques are fundamental to the design of wireless communication systems, since many crucial components of such systems depend on the choice of the multiple access technique [1]. For example, the use of orthogonal frequency-division multiple access (OFDMA) simplifies the physical layer design, where complicated channel estimation and equalization, which are mandatory for non-OFDMA systems for combating frequency selective fading, are no longer needed. The use of OFDMA has also revolutionized the design of the upper layers of a communication system, e.g., sophisticated scheduling and resource allocation schemes have been introduced by exploiting the features of OFDMA.

Because of the importance of multiple access, there has been an ongoing quest during the past decade to develop next generation multiple access (NGMA). Among the potential candidates for NGMA, non-orthogonal multiple access (NOMA) has received considerable attention from both the industrial and academic research communities, and has been highlighted in the recently published International Mobile Telecommunications (IMT)-2030 Framework as follows: "for multiple access, technologies including NOMA and grant-free multiple access are expected to be considered to meet future requirements" [2]. In particular, Ref. [3] on NOMA has demonstrated its significant potential for supporting the key usage scenarios of IMT-2030, namely massive communication, ubiquitous connectivity, integrated sensing and communication (ISAC), and hyper reliable and low-latency communication. However, there is still no consensus in the research community about how exactly NOMA assisted NGMA should be designed. The aim of this perspective is to illustrate the three key features that NOMA assisted NGMA should have, as detailed in the following.

Multi-domain utilization. Conventional NOMA and orthogonal multiple access (OMA) have been designed by focusing on the efficient use of the bandwidth resource blocks available in a single domain, e.g., time division multiple access (TDMA) relies on orthogonal resource blocks in the time domain, i.e., time slots, and many forms of NOMA are designed to utilize the power domain for multiple access. There have been some new forms of NOMA that can realize partial multi-domain utilization, as illustrated in Figures 1(a) and (c). For example, Figure 1(a) shows a type of NOMA that utilizes both the power and time domains, by first dividing users into multiple non-overlapping NOMA groups and then serving different groups in different time slots. One drawback of this type of NOMA is that each user has access to a single time slot only, and hence the overall system performance can be improved with full multi-domain utilization, i.e., each user is provided access to multiple time slots as shown in Figure 1(b). The concept of multi-domain utilization can be generalized by including the spatial domain, as shown in Figures 1(c) and (d). In particular, a typical combination of NOMA, TDMA, and space division multiple access (SDMA) is shown in Figure 1(c). The resources from the space, power, and time domains can be used in a more flexible and efficient manner via full multi-domain utilization as shown in Figure 1(d), where each user is not constrained to a single time slot.

Multi-mode compatibility. Ideally, NGMA should be compatible with different existing multiple access techniques, so that dynamic coexistence between different multiple access modes can be supported. One way to support this multi-mode compatibility is to implement NOMA as an add-on to OMA, which leads to the concept of hybrid NOMA [4]. The key idea of hybrid NOMA is illustrated in Figures 1(e) and (f). In particular, Figure 1(e) shows an example of TDMA, where four users are served in four time slots individually. Figure 1(f) shows a hybrid NOMA scheme that enables the dynamic implementation of OMA and NOMA. In particular, both users 1 and 3 adopt OMA and use a single time slot each, whereas users 2 and 4 adopt NOMA. In other words, hybrid NOMA can realize a harmonic and efficient integration of different multiple access techniques, where a user can freely choose a multiple access technique based on its own capability and quality of service requirements. We note that multi-mode compatibility also has the benefit that it can be deployed in existing OMA based networks, without the requirement of changing the current standards.

Multi-dimensional optimality. The aforementioned two features lead to this new feature from the optimization per-

* Corresponding author (email: zhiguo.ding@manchester.ac.uk)

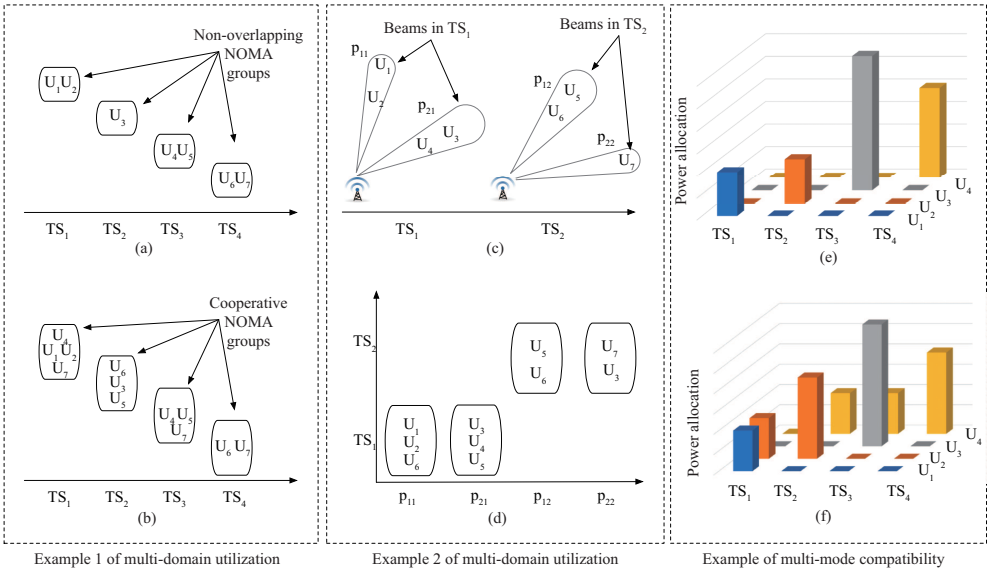

Figure 1 (Color online) Illustrations of the features of multi-domain utilization and multi-mode compatibility. (a) Partially using the power and time domains; (b) fully using the power and time domains; (c) partially using the spatial, power, and time domains; (d) fully using the spatial, power, and time domains; (e) single-mode: TDMA; (f) multi-mode: hybrid NOMA.

spective. For example, single-dimensional optimization is used to carry out resource allocation for the conventional designs shown in Figures 1(a) and (c), i.e., the system optimization is carried out with the constraint that each user has access to a single time slot only. On the other hand, the multi-domain utilization schemes shown in Figures 1(b) and (c) rely on multi-dimensional optimization, e.g., the system optimization is carried out without the constraint that each user can use only a single time slot. From the optimization perspective, an optimization problem with fewer constraints yields better performance, which is the reason why the multi-dimensional optimization schemes are expected to yield better performance than single-dimensional ones. We note that multi-dimensional optimization problems are more challenging to solve than single-dimensional optimization problems, as illustrated in the following example. For resource allocation carried out within a single dimension, i.e., a single time slot, a simple example of NOMA resource allocation is provided as follows:

Problem 1.

$$\min_{P_i\geqslant 0} f_0(P_1,P_2) \ \text{ s.t. } \ \log\left(1+\frac{P_1}{P_2+1}\right)\geqslant R, \qquad (1)$$

where $f_0(P_1,P_2)$ is an arbitrary resource allocation objective function, R is the target data rate, P_i is user i's transmit power, and $\log(1+\frac{P_1}{P_2+1})$ is user 1's data rate if user 2's signal is treated as additive white Gaussian noise. Problem 1 can be straightforwardly solved since the constraint $\log(1+\frac{P_1}{P_2+1})\geqslant R$ can be converted to an affine form: $P_1\geqslant (P_2+1)(e^R-1)$. For multi-dimensional optimization, we assume instead that user 1 can use one additional time slot with the transmit power denoted by P_0. The corresponding optimization problem can be formulated as follows:

Problem 2.

$$\min_{P_i\geqslant 0} f_0(P_1,P_2,P_0) \ \text{ s.t. } \ \log\left(1+\frac{P_1}{P_2+1}\right)+\log(1+P_0)\geqslant R. \qquad (2)$$

Comparing Problems 2 and 1, we observe that the difference between the two optimization problems seems trivial, where the simple term $\log(1+P_0)$ is added to the constraint of Problem 2. However, solving Problem 2 is significantly more challenging than solving Problem 1, since it is not a trivial task to convert constraint (2) to an affine or convex form.

Open problems for future research directions. (1) Ambient Internet of Things (IoT) with zero-energy devices. Compared to the conventional multiple access schemes shown in Figures 1(a) and (c), NGMA with the three aforementioned features requires each user to carry out more transmissions. For energy constrained communication networks, a device that carries out many transmissions can suffer a short battery lifespan. A promising solution for this energy consumption issue is to apply ambient IoT with zero-energy devices, where backscatter communication (BackCom) can be applied to ensure that the NOMA transmission is carried out in a battery-less manner. This ambient backscatter feature may be particularly useful for NGMA, as explained in the following. Take the four-user scenario shown in Figures 1(e) and (f) as an example. Assume that TDMA has been deployed in a legacy network. Hybrid NOMA can be implemented in an ambient manner, where NOMA transmission is powered by the existing transmission in the legacy network. For example, user 2 wants to carry out a NOMA transmission in the first time slot, as shown in Figure 1(f). To avoid draining its own battery by the NOMA transmission, user 2 can reflect and modulate the signal sent by user 1, i.e., user 1's signal is treated as a carrier signal by user 2. Therefore, from the communication perspective, the use of BackCom can effectively enhance the energy and spectrum cooperation among the users in NGMA networks; however, from the optimization perspective, the system optimization becomes more challenging, e.g., in addition to the users' transmit powers, the users' reflection coefficients also need to be optimized, and these system parameters are strongly coupled.

(2) Near-field communications. In addition to NOMA, SDMA has also been envisioned to be a key component of NGMA, since SDMA can efficiently utilize the spatial degrees of freedom offered by multiple-input multiple-output

(MIMO) systems. For conventional far-field MIMO systems, when the users' distances are larger than the Rayleigh distance, the combination of NOMA and SDMA is an obvious solution, as explained in the following. For far-field users, the plane-wave channel model can be used, which means that a user's far-field channel vector is solely parametrized by its angle of departure/arrival. This is the reason why a far-field beamformer is expected to cover a cone-shaped area, as illustrated in Figure 1(c). Any two users that fall within the same cone-shaped area will have highly correlated channel vectors, and hence can be served by a single beamformer via NOMA. However, if the millimeter-wave (mmWave) or terahertz (THz) bands are used, extreme massive MIMO can be implemented, which means that the Rayleigh distance becomes quite large, e.g., with a 513-element uniform linear array (ULA) and a carrier frequency of 30 GHz, the Rayleigh distance is around 1400 m [5]. For users that are located in the near-field region, the spherical-wave channel model needs to be used, which results in the so-called beam-focusing phenomenon. In particular, a user's channel vector is parametrized not only by its angle of departure/arrival but also by its distance from the base station. Therefore, two near-field users' channel vectors might not be highly correlated, which makes the implementation of the beam sharing concept shown in Figure 1(c) difficult. One recent study shows that the resolution of near-field beamforming, which measures the correlation between the users' channel vectors, is far from perfect [6]. For example, for users that are moderately close to the base station, e.g., their distances from the base station are on the order of the Rayleigh distance, the users' channel vectors can still be highly correlated, and hence the concept of beam sharing is still applicable to these near-field users. In addition, NOMA assisted NGMA can also play an important role in enabling the coexistence of near-field and far-field communications.

(3) Exploiting heterogeneous channel conditions. Unlike OMA, exploiting the users' heterogeneous channel conditions is inherent in the design of NOMA. Take two-user power-domain NOMA as an example, where the user with the weaker channel conditions is allocated more transmit power. Actually, the users' heterogeneous channel conditions are key to the performance gain of power-domain NOMA over OMA, i.e., the performance gap between the two multiple access schemes is zero if the users have the same channel conditions. However, for NOMA assisted NGMA, it is challenging to exploit the users' heterogeneous channel conditions, due to its multi-dimensional nature. For example, a typical sum-rate maximization problem for NOMA assisted NGMA is given as follows:

Problem 3.

$$\max \quad \sum_{n=1}^{N}\sum_{i=1}^{M} R_{i,n} \quad \text{s.t.} \quad \sum_{n=1}^{N}\sum_{i=1}^{M} P_{i,n} \leqslant P_{\text{total}},$$

where $R_{i,n}$ denotes user i's data rate in time slot n, $P_{i,n}$ denotes the user's corresponding transmit power, P_{total} denotes the overall power budget, N denotes the total number of time slots, and M denotes the number of users. The challenge in solving Problem 3 is the availability of the users' channel state information (CSI), which is the reason why the users' CSI was assumed to be constant in [4]. This assumption is valid only for the case of low-mobility users. A promising solution is to apply ISAC, where the users' CSI can be estimated and predicted for carrying out dynamic resource allocation.

(4) Dynamic long-term system optimization. As discussed previously, NOMA assisted NGMA needs to be designed by fully utilizing multiple domains. Problem 3 shows a typical example of multi-time-slot optimization. For the case with a finite number of time slots, this optimization problem can be solved by applying conventional convex optimization tools. However, for the case with $N \to \infty$, Problem 3 becomes challenging to solve. We note that such long-term system optimization of NGMA is important since NGMA is expected to enable a user to dynamically switch between different multiple access modes over time. One promising solution to achieve dynamic long-term system optimization is to apply reinforcement learning by modeling the resource allocation as a Markov decision process, i.e., Problem 3 is recast as follows:

Problem 4.

$$\max \quad \mathcal{E}\left\{ \sum_{n=1}^{\infty} \gamma^{n-1} \sum_{i=1}^{M} R_{i,n} \middle| \pi_{P_{i,n}}, s_0 \right\},$$

where γ denotes a discount rate parameter, $\pi_{P_{i,n}}$ denotes a policy which is a set of sequential decisions for $P_{i,n}$, the conditional expectation term $\mathcal{E}\{\cdot\}$ denotes the expected value of the discounted cumulative sum of the system throughput if policy $\pi_{P_{i,n}}$ is used, and s_0 denotes the initial network state, e.g., how many bits each user needs to deliver, the users' transmit power budgets, and the users' initial CSI. We note that this reinforcement learning approach could also be useful to efficiently exploit the users' dynamic channel conditions without the global CSI assumption, since reinforcement learning is robust to changes in the environment [7].

Conclusion. NOMA assisted NGMA has been envisioned in the recently published IMT-2030 Framework. This perspective has outlined three important features of NOMA assisted NGMA, namely multi-domain utilization, multi-mode compatibility, and multi-dimensional optimality, where important directions for future research into the design of NOMA assisted NGMA have also been discussed.

Acknowledgements Zhiguo DING was supported by UK EPSRC (Grant Nos. EP/W034522/1, H2020-MSCA-RISE-2020, 101006411). Robert SCHOBER was supported by German Research Foundation (DFG) under Project SFB 1483 (Project-ID 442419336 Empkins) and BMBF under the Program of "Souveran. Digital. Vernetzt." Joint Project 6G-RIC (Project-ID 16KISK023). Pingzhi FAN was supported by National Natural Science Foundation of China (Grant No. 62020106001). H. Vincent POOR was supported by U.S National Science Foundation (Grant Nos. CNS-2128448, ECCS-2335876).

References

1 You X H, Wang C X, Huang J, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. Sci China Inf Sci, 2021, 64: 110301
2 ITU. Framework and Overall Objectives of the Future Development of IMT for 2030 and Beyond. Technical Report ITU-R M.2160-0, 2023
3 Islam S M R, Avazov N, Dobre O A, et al. Power-domain non-orthogonal multiple access (NOMA) in 5G systems: potentials and challenges. IEEE Commun Surv Tut, 2017, 19: 721–742
4 Ding Z G, R Schober, Poor H V. Design of downlink hybrid NOMA transmission. 2024. ArXiv:2401.16965
5 Zhu J, Wan Z, Dai L, et al. Electromagnetic information theory: fundamentals, modeling, applications, and open problems. IEEE Wireless Commun, 2024. doi: 10.1109/MWC.019.2200602
6 Ding Z. Resolution of near-field beamforming and its impact on NOMA. IEEE Wireless Commun Lett, 2024, 13: 456–460
7 Sutton R S, Barto A G. Reinforcement Learning: An Introduction. Cambridge: MIT Press, 1998