**On the Compatibility Feature of Non-Orthogonal Multiple Access**

*Zhiguo Ding1, Yongxu Zhu2, and Yan Chen3*

*1Lancaster University, UK, 2Loughborough University, UK, 3Huawei Technologies, Shanghai*

1. **Abstract**

Non-orthogonal multiple access (NoMA) has recently emerged not only as a new design of multiple access techniques in cellular networks, but also as a general principle of network architecture for applications beyond cellular systems. This paper is to focus on the compatibility feature of NoMA, and examine how it can be used with other multiple access techniques, advanced physical layer technologies, as well as other communication systems other than cellular networks.

1. **Introduction**

Non-orthogonal multiple access (NoMA) has been recognized as a paradigm shift for the design of multiple access techniques in future wireless networks [1-3]. The essential idea of NoMA is to encourage radio resource sharing, which means that multiple users are served at the same frequency resource blocks and at the same time, instead of allowing a single user to solely occupy one block for a time as in orthogonal multiple access (OMA). It is worth pointing out that the NoMA principle is not new to wireless communications, and it is based on many existing ideas employing non-orthogonality. For example, the key components of NoMA systems, such as superposition coding, successive interference cancellation (SIC) and message passing algorithms, have been previously used for multi-user detection. However, intentionally multiplexing different users at the same time/frequency/code/spatial-direction has not been used in the previous generations of multiple access designs.

This paper is to focus on the compatibility feature of NoMA, which can be used to partially answer the following question: why the NoMA concept can gain huge popularity in the industrial and academic research community in such a short term. In particular, we will illustrate the compatibility of the NoMA principle from three aspects. Firstly, NoMA is compatible to all the other orthogonal multiple access techniques, and hybrid NoMA can be implemented in a straightforward manner without changing the fundamental blocks of other multiple access schemes while bringing additional benefits in improving the multi-user network performance. Secondly, NoMA provides a general network architecture, in which advanced communication techniques, such as flexible duplex, millimeter-wave (mmWave) transmission, multiple-input multiple-output (MIMO), etc., can be straightforwardly used. Finally, we show that the NoMA principle is not just applicable to cellular networks, but also to many other important communication systems.

1. **A New Era of Hybrid Multiple Access**

Hybrid multiple access is not new to cellular networks at all, and it has been employed in early generations of mobile networks. For example, GSM relies on the combination of time division multiple access (TDMA) and frequency division multiple access (FDMA). Particularly, the principle of TDMA is used to divide one GSM frame into 8 time slots, and hence 8 users can be supported at the same carrier frequency, which obviously is not sufficient to support a large number of users and motivates the use of FDMA in GSM. While in LTE, OFDMA is applied as another example of OMA where the time and frequency resources are gridded into orthogonal lattices and different users are scheduled on non-overlapping lattices exclusively. The future generation wireless networks are also expected to continue using such hybrid multiple access, and NoMA is envisioned to play an important role due to its superior compatibility. Particularly, NoMA can be used to improve spectral efficiency without any need for altering the fundamental resource blocks of other multiple access principles, where the integration of NoMA can be viewed as a smooth software upgrade.

While the principle of NoMA can be integrated with other multiple access techniques, there are various solutions to how such an integration can be done. In this paper, two popular solutions are introduced. *The first type of hybrid NoMA* is that the principle of NoMA, spectrum sharing, is implemented on each orthogonal resource element individually and separately. If the orthogonal resource element is an OFDM subcarrier within a given time unit, this type of hybrid multiple access has also been termed single-tone NoMA. The benefit of single-tone NoMA is simple in concept as well as the transceiver design, since the implementation of NoMA on one resource element is independent to those on the others, i.e., there is no need for joint NoMA encoding or decoding across different resource elements. The most well-known form of single-tone NoMA is power-domain NoMA. Particularly, power-domain NoMA invites multiple users to share the same resource element simultaneously, where the power domain is used for multiplexing, e.g., superposition coding is used at the downlink transmitter by allocating different power levels to users, and SIC is used at the receivers to remove co-channel interference. Cognitive radio inspired NoMA (CR-NoMA) is another important form of single-tone NoMA. The key difference between power-domain NoMA and CR-NoMA is that CR-NoMA recognizes the difference between users’ quality of service (QoS) requirements, in addition to the users’ channel difference. For example, one of the key features of future wireless networks is the heterogeneous traffic pattern, and such diverse QoS feature becomes more obvious after Internet of Things (IoT) is integrated with cellular networks. Compared to power-domain NoMA , CR-NoMA offers two advantages. One is that the principle of NoMA can be still implemented even if users’ channel conditions are similar, since users are ordered accordingly their QoS targets. The other is that users’ QoS requirements can be strictly guaranteed.

*The second type of hybrid NoMA* is to jointly implement the principle of NoMA across multiple orthogonal resource elements, which is thus referred as multi-tone NoMA. In such multi-tone NoMA, some types of joint coding across the multiple resource elements can be introduced to further improve NoMA performance. For instance, block based sequence spreading, scrambling, interleaving, or multi-tone joint modulation mapping can be applied. Joint decoding across multiple resource elements and multiple users is thus needed for the implementation of multi-tone NoMA. Compared to single-tone NoMA, multi-tone NoMA offers better reception reliability and throughputs, but may suffer from increased design complexity at both transmitter and receiver sides. In another aspect, because of the joint design over multiple resource elements, the performance of multi-tone NoMA is further impacted by the way of resource elements mapping or subcarrier allocation, which can also be taken as one dimension to be optimized. Currently, seeking the optimal solution for resource allocation with low complexity for multi-tone NoMA has become an important ongoing research direction.

Despite such increased design complexity, multi-tone NoMA has attracted a lot of attention from the industry, and many recently proposed industrial forms of NoMA are based on multi-tone NoMA. For example, uplink sparse code multiple access (SCMA) is implemented by asking each user jointly encodes its messages sent on multiple subcarriers, and using the message passing algorithm (MPA) at the base station for joint decoding. Regular SCMA has a strict constraint that each user can be allocated the same number of subcarriers, whereas irregular SCMA as well as pattern division multiple access (PDMA) use more relaxed requirements to the number of subcarriers allocated to each user. It is worth pointing out that many industrial forms of multi-tone NoMA are based on the open loop concept, i.e., users’ channel information is not used for subcarrier allocation. While this open loop approach reduces the system complexity, the dynamic nature of users’ channel conditions has not been used, which means that the performance of these industrial forms of multi-tone NoMA can be further improved by exploiting the users’ channel information.

1. **Compatibility of NoMA with Other Advanced Physical Layer Designs**

Extensive existing studies have demonstrated that the principle of NoMA is compatible not only to other types of multiple access, but also to advanced physical layer techniques to be used in the future wireless networks. Some of these examples are provided in the following:

*mmWave-NoMA:* Both mmWave and NoMA have been recognized as key techniques to combat the spectrum crunch, i.e., there are not sufficient bandwidth for communications, although the solutions provided by the two techniques are different. A common question from the research community is why to use NoMA for mmWave networks when there is plenty of bandwidth available at mmWave bands. This question can be answered by using the following example. In 1990s, when a movie is stored in a computer by using the VCD (MPEG-1) format, the size of such a file is around 200-500MB. When this was replaced by the DVD format, the size of a movie file is expanded to 2-3GB, and the size of a typical Blu-Ray file can be 20GB. Human become more and more demanding to the resolution and details, which means that the amount of information to be sent in the future wireless networks will also become larger and larger. Therefore, the gains we get from mmWave bands can soon hit its ceiling, and how to efficiently use the bandwidth from the mmWave bands will become a critical issue, which motivates the use of NoMA in mmWave networks.

Furthermore, existing studies of mmWave-NoMA have revealed that mmWave transmission exhibits some features which are ideal for the application of NoMA. For example, users in mmWave networks can have strongly correlated channels, even if the antennas of these users are separated much larger than half of the signal wavelength. While such correlation has been conventionally recognized as a harmful effect, the use of the quasi-degradation criterion reveals that this correlation results in an ideal situation for the application of NoMA. Another example is hardware impairments and limitations, e.g., the use of finite resolution analog beamforming, can also bring the opportunity for the integration of NoMA in mmWave networks.

*MIMO-NoMA:* During the last two decades, MIMO has been continuously in the spotlight of the communication research and industrial activities, mainly due to its superior spectral efficiency, i.e., high data rates can be supported without using extra spectrum bandwidth but by exploring the spatial domain. At a certain stage of the development of NoMA, there was confusion about the difference between MIMO and NoMA. The reason for this confusion is that using MIMO, we can also accommodate multiple users at the same spectrum at the same time, and hence yield the same non-orthogonality as NoMA. Actually, many convectional MIMO techniques aim to use the spatial domain and create spatially orthogonal channels between users, in order to avoid co-channel interference. Zero forcing and singular-value-decomposition based designs are typical examples to illustrate this orthogonal principle. On the other hand, the use of NoMA is to assume that multiple users share the same orthogonal resource unit, where one spatially orthogonal channel is just another example of such a resource unit. Or in other words, conventional MIMO allows users to use the same bandwidth, but tries to create multiple orthogonal spatial directions to differentiate multiple users, whereas NoMA further supports multiple users to share the same spatially orthogonal direction.

In the context of massive MIMO, there was a concern about the feasibility for the implementation of NoMA. The rationale behind this concern is that the quasi-degradation criterion reveals that NoMA is not preferable if users’ channel vectors are orthogonal to each other, but in massive MIMO, users’ channel vectors are asymptotically orthogonal. However, some existing studies have demonstrated that the use of NoMA is still important to massive MIMO, where the reason is that users’ channels are not completely orthogonal in a practical scenario, because of channel correlation. For example, when implementing massive MIMO at a base station, most likely this base station will be amounted at a top of a high building, without many scatters around. As a result, users from one room in this building can have highly correlated channels, instead of orthogonal channels. As discussed in the mmWave-NoMA part, the correlation among users’ channels does facilitate the implementation of NoMA. Moreover, it is very costly to get accurate channel state information for massive MIMO scenarios. In the case that the channel state information is not perfect due to limited feedback quantization, channel measurement latency, or user mobility, NoMA can help improve the system performance which will otherwise be degraded significantly.

*Cooperative NoMA:* The importance of cooperative diversity can be easily spotted by the fact that the paper by Laneman, Tse and Wornell has already attracted 12000+ citations, probably one of the most cited papers in the last two decades in communications. The use of cooperative transmission is important to NoMA since users with poor channel conditions in current NoMA can potentially suffer some performance loss, compared to the case with OMA. By using cooperative NoMA, the reception reliability of these users can be improved. Most existing designs of cooperative NoMA can be grouped into two categories. One category is to employ NoMA users with strong channel conditions as relays to help the other users, which is also known as user cooperation. The benefit of such cooperative NoMA is that the redundant structure of NoMA can be efficiently exploited. Particularly, these so-call strong users need to decode the messages to the users with poor channel conditions in order to decode their own information, and hence they are natural relays to help those weak users.

The use of dedicated relays is another important category of cooperative NoMA. In many communication scenarios, the number of mobile devices is large, but many of them are not active in transmitting or receiving. Therefore, these idle users can be used as relays to help the active users, and the number of such relays can be quite large in practice, which is the advantage of the second category of cooperative NoMA. Given the existence of multiple relays, distributed beamforming can be designed in order to efficiently utilize the spatial degrees of freedom offered by the dedicated relays, but the system overhead caused by the coordination among the relays needs to be carefully suppressed, particularly for the scenario with a large number of relays. A low-complexity alternative to distributed beamforming is relay selection, and recent studies reveal an interesting fact that relay selection in cooperative NoMA can be fundamentally different to that in conventional cooperative networks. For example, the max-min criterion which has been shown optimal in conventional cooperative networks is no longer optimal in cooperative NoMA. This explains why relay selection becomes a quite popular area in cooperative NoMA.

*Network NoMA:* Network MIMO has recently received a lot of attention, since the boundaries of cells are removed and base stations from different cells are encouraged to cooperate each other. There have been different forms of network MIMO, from distributed CoMP to jointly scheduled C-RAN. While the benefit of the NoMA principle in a single cell setup has been well recognized, its benefit to the multi-cell scenarios, such as CoMP and C-RAN, has not been fully exploited and investigated. Hence recently, a lot of efforts of the NoMA research community have been has been devoted to thoroughly examine the impact of NoMA on multi-cell scenarios, and the resulting novel designs of NoMA can be viewed as special cases of network NoMA, and these network NoMA designs clearly demonstrate that it is beneficial to use the NoMA principle in such network MIMO scenarios [4]. In particular, given its non-orthogonal nature, NoMA is expected to improve the CoMP transmission by relaxing the requirement of accurate time alignment between different transmit points and joint channel state information for precise beamforming, which prohibit the boom of CoMP applications in practical systems.

Without loss of general, take CoMP as an example, which is to ask multiple base stations to jointly serve a user which is at the cell boundaries. While such a design indeed helps the edge user, these base stations have to serve this user solely for a given bandwidth and time slot, which reduces the spectral efficiency since this user has poor connections to the base stations. After network NoMA is used, each base station can serve a near user while performing CoMP and helping the edge user. As a result, the overall system throughput as well as connectivity can be significantly improved. Advanced power allocation policies can be used to ensure that those near users are admitted without sacrificing the performance of the edge user. In the heterogeneous networks, the NoMA principle has also been shown to be very useful to improve the spectral efficiency as well as coverage, where not only more users can be served in each tier but also the cooperation among different tiers can be enabled.

1. **Integrating NoMA into Systems Beyond Cellular Communications**

The superior compatibility of the NoMA principle can be also demonstrated by the fact that NoMA has found a lot of applications beyond cellular communications, as illustrated in the following:

*Wi-Fi Networks:* While the concept of NoMA has been investigated for cellular systems, it can be straightforwardly applied to the next generation Wi-Fi systems. Conventional Wi-Fi networks still rely on orthogonal resource allocation. This leads to a difficult situation that some users cannot be admitted after all the limited orthogonal resource blocks are taken by other users. By applying the NoMA principle, more users can be simultaneously admitted, which is particularly important for the deployment of Wi-Fi in crowded areas, such as airports or sport stadiums.

*VLC Systems:* VLC has been recognized as an efficient method for the last mile connection in future communication networks. However, one disadvantage of VLC is that the use of narrow-band modulation and non-coherent detection limits the number of users which can be supported. As a result, the NoMA principle is naturally compatible to VLC, and its application ensures that VLC can be used not only for small-scale smart homes, but also large-scale scenarios, such as lecture halls.

*TV Broadcasting:* Digital TV broadcasting is another important application of NoMA. It is worth pointing out that the concept of NoMA has already been included into the next generation digital TV standard (ATSC 3.0), where it is termed Layered Division Multiplex (LDM). Particularly, a TV station will integrate several layers of video streams with different QoS requirements, and this superposition will be broadcasted to users. Each user will decode certain layers of the video streams according to its channel conditions.

*Wireless Caching:* The key idea of wireless caching is to proactively push content files to local caching infrastructure, before they are requested. As a consequence, users can fetch these files from their local caching infrastructure, without being directly served by the network controller. Existing studies have demonstrated that the NoMA concept not only helps to push the content files to local caching infrastructure timely and reliably, but also improves the spectral efficiency of content delivery from the caching infrastructure to the users.

*Internet of Things (IoT):* One key feature of IoT is the diverse traffic patterns of IoT devices. Particularly, some devices have demanding bandwidth requirements, e.g., environmental mentioning cameras deliver high-resolution images/videos, whereas the others need to be served with low data rates but timely, e.g., vehicles receive incident warning messages in intelligent transportation systems. NoMA can be naturally applied to handle such a challenging situation, by integrating devices with heterogeneous QoS requirements at the same bandwidth. Furthermore, recent studies have also demonstrated that the combination of encoding with finite block length and NoMA can be a promising solution for supporting IoT as well as ultra reliability and low latency communications (uRLLC). In particular, for contention based grant-free transmission, NoMA is the key solution to enable reliable communications while supporting massive connectivity.

1. **Conclusions**

The aim of this paper is to focus on the compatibility feature of NoMA. In particular, we have first examined how NoMA can be combined with other multiple access techniques. Then the compatibility between NoMA and various advanced physical layer techniques has been illustrated, and the applications of NoMA to communication systems beyond cellular networks have also been discussed.

**References**

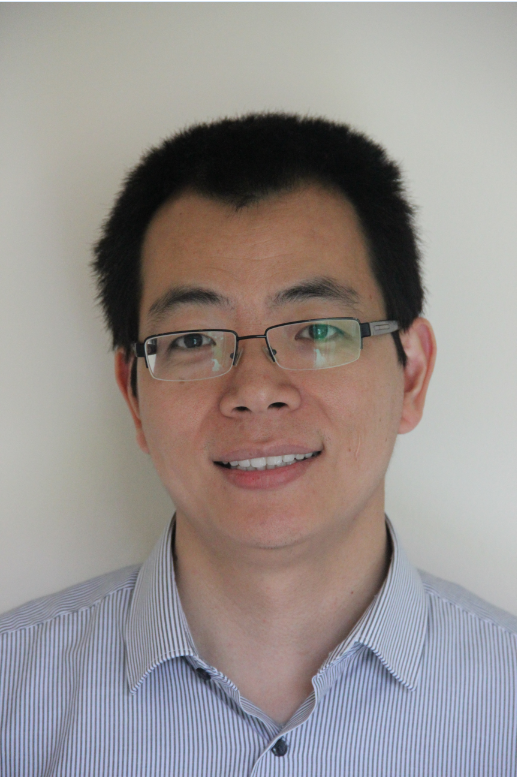
[1] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan and V. K. Bhargava, "A Survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends," in IEEE J. Sel. Areas Commun., vol. 35, no. 10, pp. 2181-2195, Oct. 2017.

[2] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C-L. I and H. V. Poor, “Application of Non-orthogonal Multiple Access in LTE and 5G

Networks”, IEEE Commun. Mag., 2017.

[3] L. Song, Y. Li, Z. Ding, and H. V. Poor, “Resource management in nonorthogonal multiple access networks for 5G and beyond”, IEEE Networks, 2017.

[4] Usa Vilaipornsawai, Hosein Nikopour, Alireza Bayesteh, and Jianglei Ma, “SCMA for Open-Loop Joint Transmission CoMP,” in Proc. IEEE VTC fall, 2015.

**Dr. Zhiguo Ding** is currently a Chair Professor in Signal Processing for Communications at School of Computing and Communications, Lancaster University, the UK.

**Dr. Yongxu Zhu** is currently a research associate at the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, the UK.

**Dr. Yan Chen** is currently with Huawei Technologies Co., Ltd., Shanghai, as a Technical Leader on Huawei 5G air interface design.