J. Indian Inst. Sci.
A Multidisciplinary Reviews Journal
ISSN: 0970-4140 Coden-JIISAD

REVIEW ARTICLE

Non-orthogonal Multiple Access: An Enabler for Massive Connectivity

Vimal Bhatia[1], *Pragya Swami*[1*], *Sanjeev Sharma*[2] *and Rangeet Mitra*[3]

Abstract | Two of the most challenging goals for the fifth generation (5G) and beyond communication systems are massive connectivity and higher capacity. The use of traditional orthogonal multiple access techniques limits the number of users that can be served using available resources due to orthogonality constraint. Moreover, the available resources may not be utilized effectively by alloted users thereby resulting in inefficiency and user unfairness. This imposes a severe drawback in cases where the number of users to be served are high, like in the Internet of Things or ultra-dense 5G networks. Hence, introducing non-orthogonality to multiple access scheme is advocated as a superior methodology to serve multiple users simultaneously, thereby achieving multi-fold enhancement in connectivity. In scenarios with massive number of users, non-orthogonal multiple access scheme (NOMA) increases the number of active connections by superimposing signals of multiple users on the same resource block, thereby also utilizing the available resources efficiently. This article presents an overview of the integration of NOMA with several other leading technologies for 5G and beyond networks to enhance the connectivity.

Keywords: *Non-orthogonal multiple access, 5G communication, Millimeter wave communication, HetNets, UDN, Cooperative communication, SWIPT, VLC, Massive MIMO, CR networks, MEC, UAV, Full duplexing*

1 Introduction

Higher capacity and larger number of active users (connected devices) than the current fourth generation (4G) network are the prime objectives for future fifth generation (5G) technology[1]. However, the present use of *orthogonal multiple access (OMA)* schemes at physical layer for access methods restricts the number of users that can be simultaneously served using available resources. OMA, e.g., frequency division multiple access, time division multiple access, code division multiple access, and others, works by assigning orthogonal resources to users, which cannot be utilized by other users until the resources are freed by the ongoing user. This can lead to severe degradation in overall system throughput performance and can also limit the number of users that can be served at a given instance using the available resource block (RB). This can be explained using a simple example; consider a scenario when a RB is alloted to a user with poor channel gain towards base station (BS). Due to the poor channel, the user cannot effectively utilize the alloted RB, neither can another user use the same RB for its own transmission due to the orthogonal allocation scheme. This effectively wastes the resources and hence, it becomes necessary to bring advanced technologies, e.g. new multiple access (MA) scheme, to utilize the available resources efficiently. Breaking the convention of orthogonality, non-orthogonality at access network has emerged as a leading solution to

[1] *Discipline of Electrical Engineering, Indian Institute of Technology Indore, Indore, India.*
[2] *Department of Electronics Engineering, Indian Institute of Technology (BHU) Varanasi, Indore, India.*
[3] *ETS Montreal, Quebec, Canada.*
**phd1601102001@iiti.ac.in*

reduce wastage of resources, wherein the same RB is shared by multiple users at the same instant of time, henceforth increasing the number of simultaneous connections. This MA scheme which utilizes non-orthogonality is termed as **non-orthogonal multiple access (NOMA).** NOMA serves multiple users using the same resources, e.g., frequency, code, space, etc., at the same instant of time. One prime advantage of using NOMA, by virtue of its application, is that it can serve a user with a better channel condition and a user with a poorer channel condition using the same resource, thereby increasing the number of active users and additionally provides efficient utilization of available resources[2].

NOMA: NOMA serves multiple users using the same resource, e.g., frequency, code, space, etc., at the same instant of time.

Recently, various NOMA schemes have been proposed and discussed in the literature. Some of the popular NOMA schemes includes interleave division multiple access[3], power domain **non-orthogonal multiple access (PD-NOMA)** [4–6], low density spreading CDMA[7], sparse code multiple access (SCMA)[8], pattern division multiple access[9], and multi-user sharing access[10]. The PD-NOMA schemes have been discussed for long term evolution (LTE) in the third generation partnership programme (3GPP) as multi-user superposition transmission or popularly referred to as MUST[5]. The scheme was recently approved and the discussion on PD-NOMA's inclusion in the new radio (NR) has been initiated[11]. Hence, this article focus primarily on the PD-NOMA (referred as NOMA hereafter) scheme and its co-existence with other contemporary and emerging 5G techniques for a wider view of its application in the future generation networks.

PD-NOMA: As its principle approach, PD-NOMA allocates different power levels to multiple users with different channel conditions and the signals of multiple users are then superimposed and transmitted from BS to the users in downlink PD-NOMA.

The key idea behind PD-NOMA principle is multiplexing using power diversity. While studying PD-NOMA, three of the foremost terminologies that will be frequently used/referred are: channel difference between users, power splitting and **successive interference cancellation (SIC).** As its principle approach, PD-NOMA allocates different power levels to multiple users with different channel conditions and the signals of multiple users are then superimposed and transmitted from BS to the users in downlink PD-NOMA. The user with better channel condition uses SIC to decode and remove message of the user with poorer channel condition before decoding its own message. For de-multiplexing of the signals at the receiver it is favorable that difference in allocated power is sufficiently large. This implies that difference in channel condition between the users multiplexed using NOMA should also be sufficiently large. The impact of channel gain difference between the users on throughput gain achieved by NOMA is studied in[12, 13]. It must be noted that power allocation in NOMA primarily depends on difference in channel gain between the users and hence plays a vital role in the overall throughput performance of NOMA systems[13]. The basic idea of power allocation in NOMA is different from the usual notion of power control in cellular systems, however, the power control algorithm follows similar procedure[14].

SIC: SIC involves decoding and subtracting/removing message of user with better channel condition.

In PD-NOMA, power allocation is different from the conventional power splitting techniques which allocates more power to users with better channel condition. Using PD-NOMA, a weak user, i.e. a user with a poorer channel condition is given larger power. This balances the trade-off between system throughput and user fairness which is of prime importance in any wireless communication system. The traditional power splitting technique achieves better overall system throughput, however, at the cost of weaker user's throughput. Also, it is important to note that although power splitting in NOMA reduces power given to a particular user when compared with OMA, the achieved throughput gain is due to joint usage of the assigned bandwidth which otherwise is distributed amongst the users under OMA. Comparison between the throughput achieved by OMA (specifically orthogonal frequency-division multiple access (OFDMA)) and NOMA using a practical scenario of two-user is illustrated in Fig. 2, where one user equipment (UE), called as UE 1, has signal-to-noise-ratio (SNR) equal to 0dB, and second UE, termed as UE 2, has SNR equal to 20dB. Further, R_1 and R_2 denote for the throughput of UE 1 and UE 2, respectively. It can be observed from Fig. 2 that NOMA achieves an increment in throughput by 32% for UE 1 and by 48% for UE 2, in comparison to when served using OMA under same SNR values. Furthermore, the factor that makes NOMA even more popular is its compatibility with OMA receivers and transmitters; one needs only small modifications in the software and little or no hardware changes for its implementation. Hence, NOMA is advocated as an ideal enabler of massive connectivity in 5G networks and beyond.

2 NOMA and Its Application to Three Main Use-Cases of 5G

The **5G technology is categorized as three basic use-case scenarios namely enhanced mobile broadband (eMBB), massive machine type**

5G technology: The 5G technology is categorized as three basic use-case scenarios namely eMBB, mMTC and URLLC.

communications (mMTC), and ultra-reliable and low latency communications (URLLC)[16].Since, NOMA is envisioned as a prime MA scheme for the 5G technology it becomes important to discuss its application for the practical use cases of 5G technology. The expected purpose of advanced MA scheme for 5G are: achieving points on the boundary of the capacity region of the multiple access channel, aiding overloaded transmissions, and availing low latency grant free transmission. All such functionalities are fulfilled by NOMA, hence, NOMA renders itself suitable for an mMTC application which requires massive connectivity. Of course, the benefits of NOMA come with additional receiver complexity, however, the mMTC is expected to be uplink centric. In the uplink communication, the UE are not subjected to higher receiver complexity, rather complexity increases only at the network side which can be easily accommodated. Furthermore, mMTC requires only small packet transmissions. For small packet transmission in uplink scenario, UEs do not require explicit need of scheduling grant from evolved NodeB to reduce power consumption at battery operated UE and also for lowering the overhead and latency. NOMA has proven to yield better throughput than its counterpart MA scheme, OMA, in grant-free transmission. This is because, bandwidth available for users is larger in NOMA, thereby, the collision probability, when same resource is used by multiple users, is reduced. It should be noted that small packet transmission are not only limited to mMTC transmissions. Similar traffic is also present in the eMBB and URLLC applications[17]. As per the discussion on the grant-free uplink transmission, NOMA is rendered suitable for the eMBB and the URLCC application also. For above basic usage scenarios of eMBB, mMTC and URLLC, typical use cases like relay and vehicle-to-vehicle are analyzed in[18].

3 NOMA with Contemporary Emerging Technologies

Various other emerging technologies contemporary to NOMA for achieving the goals of 5G technologies are multiple-input multiple-output (MIMO) and massive MIMO, millimeter wave (mmWave) communication, cognitive radio (CR), cooperative communications, energy harvesting [also termed as simultaneous wireless information and power transfer (SWIPT)], visible light communications (VLC), mobile edge computing (MEC), etc. Few other technologies that have recently come out and have been studied in the future generation network are, full duplexing (FD) and ultra dense networks (UDN), to name a few. Applications like unmanned aerial vehicle (UAV) are also gaining popularity. The key idea of any developing technology is to enhance the system performance in the given available resources and serve as many number of users as possible. To utilize the available resources, MA schemes are vital to be studied in context of all the upcoming technologies. Owing to the flexibility of NOMA, it can be easily integrated with most of the other 5G enhancing technologies. Hence, this article explores various applications of NOMA with the available emerging technologies to meet the diverse requirements of 5G and Beyond cellular network as shown in Fig. 3.

3.1 NOMA with mmWave Communication

Large chunks of underutilized bandwidth are available at higher frequencies ranging from 30 to 300 GHz referred to as **mmWave frequency**, that can resolve congestion and also provide higher bandwidth as compared to the limited resources available for the 4G communication[19]. Although accessible bandwidth is huge at the mmWave frequencies, NOMA serves as an essential MA technique. The physical nature of the mmWave communication makes NOMA even more preferable for its environment, as shown in Fig. 3I. The mmWave communication is directional in nature owing to its large path loss which requires suitable antennas to focus the waveform in a particular direction to cope with high propagation losses. The directional property makes channel gain of users correlated. Such correlation leads to severe degradation in conventional OMA technique, however, for NOMA application, channel correlation results in boosting the overall performance[20]. NOMA involves serving multiple users using the same resource block at the same time. Hence, correlated users' channel aids in pairing/grouping users to be served simultaneously using NOMA. Such correlations degrades system performance while using OMA, since, orthogonal allocation of resource requires that correlation between the channels is minimum. Ideally, orthogonal allocation of resource requires that no correlation is present, however, this is not practically possible. Wide bandwidths available at the mmWave frequency are capable of supporting a large number of users, however, sparse channels in mmWave communication limits simultaneous connections of multiple users[19, 21]. Hence, NOMA with mmWave communication can achieve tremendous increase in connectivity especially in overcrowded areas. Furthermore, spectral efficiency is

mmWave frequency: Large chunks of underutilized bandwidth are available at higher frequencies ranging from 30 to 300 GHz referred to as mmWave frequency that can resolve congestion and also provide higher bandwidth as compared to the limited resources available for the 4G communication.

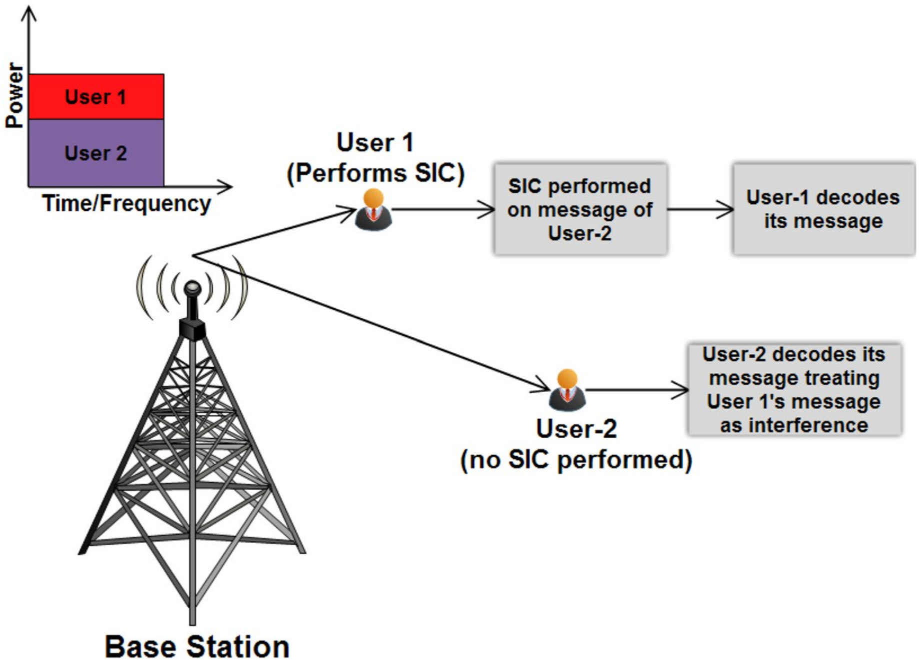

Figure 1: Basic downlink PD-NOMA architecture.

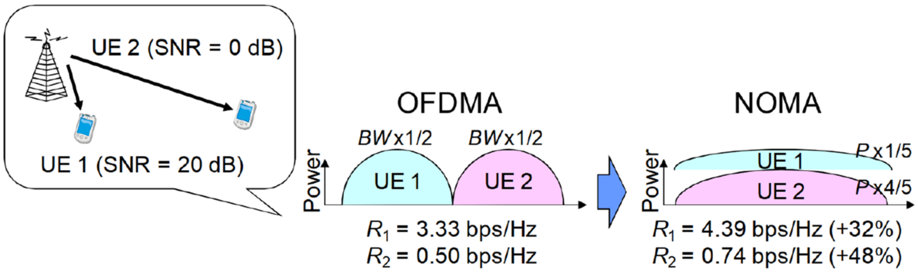

Figure 2: OMA and NOMA throughput comparison[15].

vital for the mmWave communication with the upcoming growth of virtual and augmented reality which is said to lower the radio spectrum gains obtained by the mmWave bands. Furthermore, standalone mmWave BSs are not a feasible choice for a practical system. The reason is the delay caused in initial access[22] using thin beams (a process known as beamtraining). The literature suggests that future wireless network will comprise a hybrid network with Sub-6 GHz BSs co-existing with mmWave BSs. Such hybrid network can rule out the high access delay in beamtraining[23–25].

HetNets: Heterogeneous networks refers to a network where different types of BSs co-exists to meet the requirements.

UDN: UDN refers to a wireless environment where the number of access points are larger than the number of active users.

3.2 NOMA with Heterogeneous Networks and UDNs

The term ultra dense network or **UDN** refers to a wireless environment where the number of access points (APs) are larger than the number of active users. Large number of BSs are deployed for extensive spatial reuse, which is another way to utilize spectrum efficiently[26, 27]. Future generation networks are moving towards **heterogeneous networks** (**HetNets**), a term that refers to networks where different types of BSs co-exists to meet different requirements. To fulfill various requirements, macro BSs (MBS) and the small BSs (SBS) are deployed in the same space. The MBSs are capable of providing wide coverage while the SBSs boost capacity of the network and enhance the data rates due to smaller coverage areas. NOMA integrated with such HetNets has been extensively studied in the literature as in[28, 29]. Furthermore, heterogeneous UDNs require that the deployment of the MBSs and the SBSs follow the rule as per the UDN network demands, i.e., the number of BSs needs to be larger than the

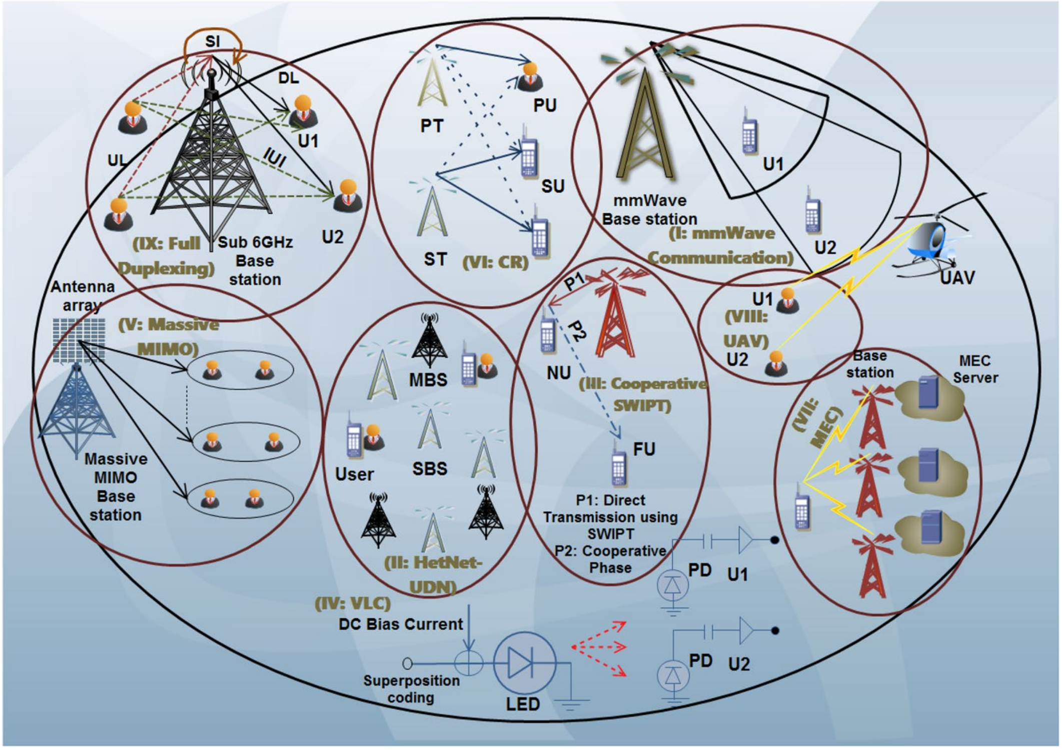

Figure 3: NOMA with various emerging technologies for 5G technology.

number of users, as shown in Fig. 1II. This implies that in a heterogeneous UDN the MBS and SBS will have a very dense deployment. Although, network densification improves capacity, however, the aggregate interference caused by interfering BSs limits the throughput performance that can be achieved. Denser the deployment of the BSs, larger is the interference from neighboring users[30]. However, on the brighter side, dense deployments ensure that each BS serves fewer users. Moreover, distance between users and BSs becomes less, thereby lowering the impact of path loss. To counter the dominant impact of increased interference of the dense networks, several innovative technologies are studied to design the architecture of heterogeneous UDNs to make the management of the network easier[31]. These are software-defined networking (SDN)[32], network function virtualization (NFV)[32], cloud computing[33], and fog computing[34]. Implementation of UDNs implies massive number of connections with high data rates. The APs are connected to a server for control and management of users' access. Applying NOMA protocols, where signals of different users are multiplexed over a single RB, the number of active connection can be further boosted. In context of UDNs, where the number of users and the number of APs are nearly similar, NOMA can be implemented by multiplexing different APs over the same RB[26, 27]. NOMA enhances the overall system throughput and the spectral efficiency of the UDNs.

However, integrating NOMA in UDN poses another challenge of severe inter-cell interference from both intra-tier, and inter-tier BSs and intra-cell or inter-user interference (IUI) caused by NOMA. Hence, a thorough study of inter-cell interference management is required.

3.3 NOMA with Cooperative Communications Using SWIPT

The coverage for the cell edge users, i.e., users located at the cell edge, away from the BSs, can be improved by using **Cooperative communication** using relays which receive signal from the BS and retransmit it to the cell edge users.Several types of relaying techniques have been studied in the existing literature[35, 36]. Combining NOMA with cooperative communication is a natural extension since, as per its principle, users receive a superimposed signal of multiple users multiplexed using NOMA. The cell center user, i.e., the user located near the cell center, applies SIC before decoding its intended message. This

Cooperative communication: Using cooperative communication reduces the attenuation suffered while the signal propagates to reach cell edge user.

implies that the cell center user has the message of the cell edge user. Therefore, the signal of cell edge user can be retransmitted by the cell center user, e.g., in scenarios when direct link does not exist between the BS and the cell edge user. The use of cooperative communication reduces the attenuation suffered while signal propagates to reach the cell edge user[37–39], since, signal is retransmitted from a node/user located closer to the cell edge user. This enhances the throughput performance of the cell edge user. However, it must be noted that UEs and relays are generally battery operated devices and retransmitting signal can drain their batteries quickly. Hence, some advanced approach is required for an energy efficient design. In this context, the concept of energy harvesting (known as SWIPT in wireless communication) strives to be of great interest for researchers. The concept of **SWIPT**, first proposed in[40], has gained attention of many in enabling the design of an energy efficient networks. NOMA applied with SWIPT for cooperative communication will not only lead to spectral efficiency but will also yield an energy efficient approach for serving the cell edge user using relaying/cooperation. The authors in[40] assumed that by using SWIPT technology, the radio signals can be used in a two-fold manner, the first use is for information transfer and the second use is for harvesting of energy. However, one of the assumptions that was not practical in[40] was that both, information transfer and energy extraction was assumed to be carried out at the same instant of time. A more practical and efficient way for SWIPT is to split it into two distinct tasks which are performed in two phases, the first for energy extraction and the second for information transmission. The energy harvested in the first phase by the relays/cooperating users can be used in the second phase for information transfer to the cell edge user, which is a very energy efficient approach. Also, as discussed in Sect. 1, in NOMA, a cell center user is served using the same resource as that allotted to the cell edge user. Thus, the performance of cell edge user will deteriorate due to power splitting and interference from the cell center user. Hence, using the cell center user to assist in relaying information to the cell edge user, as shown in Fig. 3III, compensates such sharing of resources and also increases spectral efficiency of the system[41]. Furthermore, use of SWIPT enabled cell center users rules out the possibility of battery drainage during relaying.

VLC: VLC is a shift towards the nano-meter wave range where the available bandwidth can be expanded so as to meet the spectrum demands for future generation networks.

LED: VLC comprises of light emitting diode (LED) lamps, as a source of light and as a optical transmitters. At the receiver end, photodiode-arrays are used in a typical optical attocell.

SWIPT: Using SWIPT technology, radio signals can be used in a two-fold manner, first is for the information transfer and second is for harvesting energy.

3.4 NOMA with VLC

Beyond the radio frequency spectrum, there exists visible light communication (**VLC**)[42], a widely accepted complementary technology. VLC is a shift towards the nano-meter wave range where the available bandwidth can be expanded so as to meet the spectrum demands on the future generation networks. VLC is a core part for the light fidelity (LiFi) systems which is a proposed technology for various applications, for instance, Internet of Things devices, underwater communications, vehicle-to-vehicle communication, etc. VLC comprises of light emitting diode (**LED**) lamps, as a source of light and as a optical transmitters, as shown in Fig.1IV. At the receiver end, photodiode-arrays are used in a typical optical attocell[43]. The modulation by the LEDs is done at a speed that can not be perceived by the human eye[44], therefore, making the use of LEDs two-fold. One use is for illumination and the second use is for communication. Moreover, the power required by the LEDs for signal transmission is very low as compared to the requirements in radio frequency communication. NOMA is proposed as a viable option for candidate MA scheme in VLC. The major reasons for considering NOMA in VLC are[45]: VLC poses a very small communication range, hence the channel is dominated by the line-of-sight (LOS) component which promotes accurate channel estimation. The quality of channel estimation has a vital role in determining the performance of NOMA systems, since the SIC carried out at the receiver involves decoding and removal of the messages until the desired signal is reached. SIC critically depends on the quality of the channel estimation. In case of imperfect SIC, error propagates at every step of SIC and the accumulated error becomes so high that the overall throughput performance degrades. In the early literature on random access protocols in packet radio networks, capture effect has been widely studied[46, 47]. It refers to the possibility of a stronger signal being correctly decoded by the receiver when two (or more) non-orthogonal packet signals collide simultaneously. Capture effect has been shown to improve throughput. It has also been shown that SIC capability at the receiver can be exploited to achieve even further improvement in throughput. Furthermore, two parameters other than channel gain, namely, tuning angle and field of view, give additional degrees of freedom, which can be utilized for multiplexing multiple users' signals. Although current literature has investigated the integration of NOMA in VLC, the throughput gains expected are limited due to the non-linearity posed by the LEDs[48].

As a solution, the non-linearity inherited by the LEDs can be eliminated using pre-distortion or post-distortion techniques. The authors in[49] takes into consideration the non-linearity posed by the LEDs to design a pre-distorter. For this, a modified Chebyshev-normalized least mean squares (NLMS) based pre-distortion is proposed with hybrid eigen-decomposition based precoding in a MIMO NOMA-VLC setup.

3.5 NOMA with Massive MIMO

MIMO refers to use of multiple antennas at the transmitter end, and using multiple antennas at the receiver end so that multi-path propagation can be exploited to increase the system performance[50]. MIMO has emerged as a key component and has been included in new LTE-Advanced and 5G deployments[51]. From a theoretical concept that assumed unlimited number of antennas[52] to practical implementations which is deployed in the LTE-Advanced network, with up to 64 antennas at the BS, MIMO has gained tremendous popularity. **Massive-MIMO** refers to a scenario where the BSs are equipped with a larger number of antennas as compared to the number of users served by the BSs. Application of NOMA in MIMO networks, as shown in Fig. 3V, have been well analyzed in[53, 54]. However, very few have analyzed the impact of NOMA as applied to massive MIMO. One example work applies NOMA to multi-antenna BSs for performance enhancements[55]. This study compares NOMA-MIMO setup with the conventional MIMO setup with a typical zero-forcing (ZF) massive MIMO beamforming technique and reports that NOMA does not outperform ZF scheme in every scenario. However, in situations where the number of users is very large, NOMA outperforms the ZF scheme, thus showing that NOMA has a key role to play in massive connectivity.

3.6 NOMA with CR Networks

CR is a spectrally efficient approach for sharing the spectrum available for communication. The CR networks use an adaptive and an intelligent method that detects availability of vacant spectrum (or spectrum holes) and enables concurrent transmissions to serve secondary users (SU) in the spectrum holes. Various techniques like underlay, overlay, and hybrid approaches have been proposed which decide the manner in which the spectrum sharing takes place[56]. In the underlay approach, secondary transmissions [by secondary transmitter (ST)] takes places together with primary transmission [by primary transmitter (PT)] conditioned on the limit of interference caused at primary user (PU). Using the underlay technique, transmission takes place only when spectrum holes are detected; otherwise no transmission is performed. Hybrid approach uses an intelligent combination of underlay and overlay techniques as per the system requirement. CR networks combined with NOMA can effectively increase the number of users served over an RB. Researchers have been looking into integrating CR networks with NOMA, as a possible and effective combination for 5G, as shown in Fig. 3VI. However, interference caused in the CR networks due to NOMA poses a challenge. The CR-NOMA networks are envisioned to be spectrally efficient with high system capacity. Some of the existing literature have investigated the throughput performance of NOMA integrated with CR networks as in[57, 58]. The authors in[57] discussed CR-NOMA as a special form of PD-NOMA, wherein the requirements of the SU and the PU are strictly met so that excellent system performance can be achieved. In[58], NOMA is used to boost the number of accessible SUs sharing limited and dynamic licensed spectrum holes.

3.7 NOMA with MEC

MEC is capable of meeting requirement for real-time mission-critical application such as vehicular networks, connected cars, health care, smart venues, edge video caching, and many others. MEC extends the centralized cloud computing capability to UEs/devices located near to the source of information (and hence the term ‘edge’). The key idea behind the use of MEC is that the application’s performance is much better when the related processing and tasks runs closer to the source, since it lowers the network congestion. MEC enables better utilization of mobile backhaul network, however, there still exists several challenges, such as, ultra-low latency, energy-efficient computation, etc. Latest set of applications like virtual reality, augmented reality, interactive gaming overload the existing mobile networks. For instance, virtual reality requires UEs to handle several tasks such as object recognition, vision base tracking, etc. For gaming based on virtual reality, a UE will have to carry out mixed reality and human computer interaction[59]. UEs are not so efficient in terms of handling high computational complexity tasks because they are battery operated, with only a limited power source. Intensive computations may lead to complete drainage of UE batteries which may result in not meeting task deadlines. Therefore, MEC systems operate

MIMO: MIMO refers to use of multiple antennas at the transmitter end, and using multiple antennas at the receiver end so that multi-path propagation can be exploited to increase the system performance.

Massive-MIMO: Massive-MIMO refers to a scenario where the BSs are equipped with large number of antennas as compared to the number of users served by the BS.

MEC: The key idea behind the use of MEC is that the application performance is much better when the related processing and tasks runs closer to UE, since it lowers network congestion.

CR: The CR networks uses an adaptive and an intelligent method that detects availability of vacant spectrum (called as spectrum holes) and enable concurrent transmissions in the spectrum holes.

by providing and deploying facilities for computations at edge of network. Such facilities involve APs/SBSs with MEC servers, as shown in Fig. 3VII. UEs can offload their tasks to the MEC facilities provided. Collaboration of NOMA and MEC is seen as another important communication technique for future generation wireless networks. Benefits of such collaborations have been studied in[60, 61]. In the system model proposed in[60], multiple users simultaneously offload their computation tasks to one multi-antenna BS over the same time/frequency resources for remote execution, and the BS uses SIC for information decoding. To capture the potential gains of NOMA in the context of MEC,[61] proposed an edge computing aware NOMA technique which can enjoy the benefits of uplink NOMA in reducing MEC users' uplink energy consumption.

3.8 NOMA-Aided UAV Networks

UAVs: UAVs or drones are aircraftswith no human on-board tocontrol their movements.

UAVs or drones are aircrafts with no human on-board to control their movements. Movements of UAVs are managed using a ground-based controller and communication setup between the ground-based controller and the UAV. Initial use of UAVs were meant only for military purposes, however, with time the potential of UAVs is now well-recognized in the areas of technology and manufacturing. Various applications of UAVs in civil scenarios include wildfire management, distribution of cargos, aerial photography, etc.[62]. Both industries and academia have shown keen interest in building communication systems for such applications. For instance, some industrial applications include the Google Loon project, the Facebook's delivery drone[63], and the airborne LTE services from AT&T for availing global massive connectivity. To integrate the UAVs for the communication purposes in the 5G networks, it becomes essential to explore suitable MA schemes. Since, NOMA is already a popular MA scheme for the 5G networks, combination of UAVs with NOMA, as shown in Fig. 3VIII, is a promising candidate for future wireless connectivities. The work in[64] proposed a novel framework for UAV networks with massive access capability supported by NOMA and provided performance evaluation of NOMA-enabled UAV networks by adopting stochastic geometry to model the positions of UAVs and ground users.

3.9 NOMA with FD

FD: FD is a mode of communication wherein the date are transmitted and received at the same time and over the same channel.

FD is a mode of communication wherein data is transmitted and received at the same time and over the same channel. Conventionally, data transmission and data receiving are done either on different channels, or at different time instants on the same channel, or both. However, to utilize the spectrum efficiently, simultaneous transmission and reception of data over the same channel is seen as a candidate approach[65]. FD aims to provide reliable uplink and downlink transmissions simultaneously. However, the biggest challenge is self-interference (SI), as shown in Fig. 3IX. As can be inferred from the application of NOMA and full duplexing, they both are complementary in principle. Hence, merging the two techniques can prove fruitful in improving the spectral efficiency of 5G networks[66]. Furthermore, as discussed in Sect. 3.3, cooperative NOMA can be combined with FD, see[37, 66, 67]. Together, once again, the number of connections that can be simultaneously maintained is increased.

3.10 Hybrid NOMA Techniques

Users' channel gain difference: Users' channel gain difference and their distribution around BS motivates the use of a hybrid MA (HMA) system by combining the PD-NOMA and SCMA techniques

In any practical communication system, both cell center users and cell edge users exists. Furthermore, there may be fewer cell center users than cell edge users. **Users' channel gain difference** Users' channel gain difference and their distribution around BS motivates the use of a hybrid MA (HMA) system by combining the PD-NOMA and SCMA techniques. HMA enhances spectral efficiency and accommodates a larger number of users in the system as compared to the conventional PD-NOMA or SCMA for next-generation networks[68]. In HMA, both message passing algorithm (MPA) and SIC based method are used for detection[68]. For instance, consider a downlink system with one BS and J users, such that J_1 users are cell center users (called as Group-1) and J_2 are cell edge users (called as Group-1), as shown in Fig. 4. Users within a group use the SCMA technique and groups can be distinguished using PD-NOMA techniques. Thus, users of both the groups can communicate using say orthogonal resources in the system. Therefore, spectral efficiency of the HMA-based system is higher than that of the conventional PD-NOMA and SCMA systems, thanks to the higher flexibility in terms of user pairing.

4 Future Challenges

Although a lot of literature exists in the co-existence of NOMA with other emerging technologies, many technical challenges that need to be addressed in detail still remain to be explored. These challenges include optimal user pairing for NOMA in massive MIMO setups, integrating

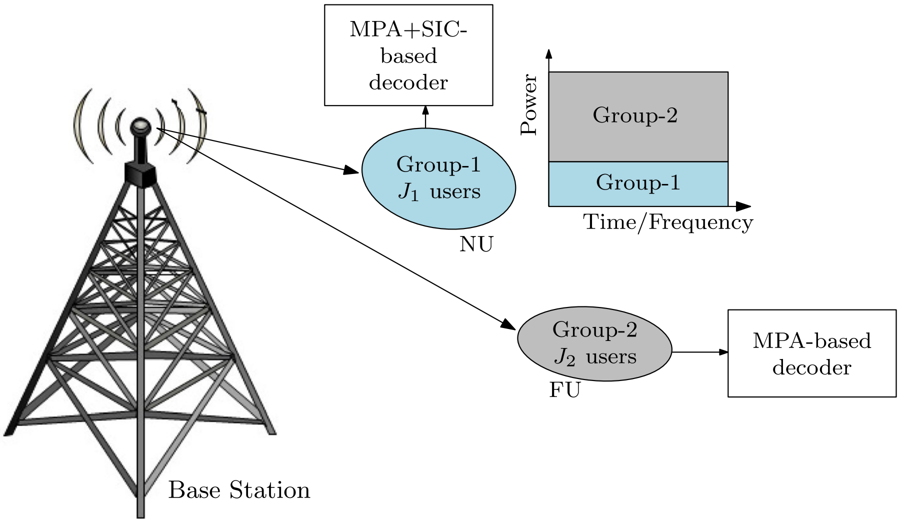

Figure 4: Hybrid multiple access method for 5G and beyond.

NOMA techniques into UAV networks, detailed study of VLC MIMO NOMA, eliminating inter-cell interference in H-UDNs with NOMA, evaluating the robustness of NOMA against synchronization errors and inter-cell interference when applied in a multi-cell scenario, and the impact of partial interference cancellation caused by imperfect channel estimation. Theoretical performance analysis for a better understanding of the impact of NOMA on MEC also needs a careful study. The responses of the research community to the above challenges will make NOMA a key enabler for massive connectivity.

Publisher's Note

Acknowledgements

This work is supported in part by the R&D project under the Visvesvaraya PhD Scheme of Ministry of Electronics and Information Technology, Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia) and Indian Institute of Technology Indore, Indore, India.

Received: 3 December 2019 Accepted: 17 March 2020

References

1. Series M (2015) IMT Vision–framework and overall objectives of the future development of IMT for 2020 and beyond. Recommendation ITU, p 2083
2. Ding Z, Fan P, Poor HV (2016) Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions. IEEE Trans Veh Technol 65(8):6010–6023
3. Ping L, Liu L, Wu K, Leung WK (2006) Interleave division multiple-access. IEEE Trans Wirel Commun 5(4):938–947
4. Higuchi K, Benjebbour A (2015) Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access. IEICE Trans Commun 98(3):403–414
5. 3GPP (2015) Study on downlink multiuser superposition transmission (MUST) for LTE (Release 13), TR36.85
6. 3GPP NTT-DOCOMO (2016) Initial views and evaluation results on non-orthogonal multiple access for NR, R1-165175
7. Hoshyar R, Wathan FP, Tafazolli R (2008) Novel low-density signature for synchronous CDMA systems over AWGN channel. IEEE Trans Signal Process 56(4):1616–1626
8. 3GPP, Huawei, HiSilicon (2016) Sparse code multiple access (SCMA) for 5G radio transmission, R1-162155
9. Dai X, Zhang Z, Bai B, Chen S, Sun S (2018) Pattern division multiple access: a new multiple access technology for 5G. IEEE Wirel Commun 25(2):54–60
10. Yuan Z, Yu G, Li W, Yuan Y, Wang X, Xu J (2016) Multi-user shared access for internet of things. In: IEEE vehicular technology conference, VTC Spring, pp 1–5
11. 3GPP, MediaTek Inc., CMCC, etc. (2016) New work item proposal: downlink multiuser superposition transmission for LTE, RP-160680

12. Swami P, Bhatia V, Vuppala S, Ratnarajah T (2017) User fairness and performance enhancement for cell edge user in NOMA-HCN with offloading. In: IEEE vehicular technology conference (VTC Spring), pp 1–5
13. Swami P, Bhatia V, Vuppala S, Ratnarajah T (2018) Joint optimization of power allocation and channel ratio for offloading in NOMA-HetNets. In: IEEE globecom workshops (GC Wkshps), pp 1–6
14. Razavi R, Dianati M, Imran MA (2017) Non-orthogonal multiple access (NOMA) for future radio access. 5G mobile communication. Springer, Berlin, pp 135–163
15. Benjebbour A, Saito Y, Kishiyama Y, Li A, Harada A, Nakamura T (2013) Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access. In: IEEE international symposium intelligent signal process. Communication system (ISPACS), pp 770–774
16. (2015) Recommendation ITU-R M.2083: IMT Vision, Framework and overall objectives of the future development of IMT for 2020 and beyond
17. 3GPP, Huawei, HiSilicon (2016) Multiple access for UL small packets transmission, R1-164036
18. 3GPP, CATT (2016) Discussion on scenarios and use cases for multiple access, R1-164246
19. Chandra K, Marcano AS, Mumtaz S, Prasad RV, Christiansen HL (2018) Unveiling capacity gains in ultradense networks: using mm-wave NOMA. IEEE Veh Technol Mag 13(2):75–83
20. Zhu L, Zhang J, Xiao Z, Cao X, Wu DO, Xia X-G (2019) Joint Tx–Rx beamforming and power allocation for 5G millimeter-wave non-orthogonal multiple access (MmWave-NOMA) networks. IEEE Trans Commun 20:20
21. Thornburg A, Bai T, Heath RW Jr (2016) Performance analysis of outdoor mmWave ad hoc networks. IEEE Trans Signal Process 64(15):4065–4079
22. Li Y, Andrews JG, Baccelli F, Novlan TD, Zhang J (2016) On the initial access design in millimeter wave cellular networks. In: IEEE Globecom Workshops (GC Wkshps), pp 1–6
23. Ghatak G, De Domenico A, Coupechoux M (2018) Coverage analysis and load balancing in HetNets with millimeter wave multi-RAT small cells. IEEE Trans Wirel Commun 17(5):3154–3169
24. Swami P, Mishra MK, Bhatia V, Ratnarajah T (2019) Outage probability of ultra high frequency and millimeter wave based HetNets with NOMA. In: IEEE international symposium on wireless communication systems (ISWCS), pp 166–170
25. Swami P, Mishra MK, Bhatia V, Ratnarajah T (2020) Performance analysis of noma enabled hybrid network with limited feedback. IEEE Trans Veh Technol. https://doi.org/10.1109/TVT.2020.2974004
26. Zhang Z, Yang G, Ma Z, Xiao M, Ding Z, Fan P (2018) Heterogeneous ultradense networks with NOMA: system architecture, coordination framework, and performance evaluation. IEEE Veh Technol Mag 13(2):110–120
27. Qin Z, Yue X, Liu Y, Ding Z, Nallanathan A (2018) User association and resource allocation in unified NOMA enabled heterogeneous ultra dense networks. IEEE Commun Mag 56(6):86–92
28. Swami P, Bhatia V, Vuppala S, Ratnarajah T (2018) A cooperation scheme for user fairness and performance enhancement in NOMA-HCN. IEEE Trans Veh Technol 67(12):11965–11978
29. Swami P, Bhatia V, Vuppala S, Ratnarajah T (2018) On user offloading in NOMA-HetNet using repulsive point process. IEEE Syst J 13(2):1409–1420
30. Nguyen VM, Kountouris M (2016) Performance limits of network densification. arXiv:1611.07790 (arXiv preprint)
31. (2017) System architecture for the 5G system, 3GPP TS 23.501
32. Granelli F, Gebremariam AA, Usman M, Cugini F, Stamati V, Alitska M, Chatzimisios P (2015) Software defined and virtualized wireless access in future wireless networks: scenarios and standards. IEEE Commun Mag 53(6):26–34
33. Checko A, Christiansen HL, Yan Y, Scolari L, Kardaras G, Berger MS, Dittmann L (2014) Cloud RAN for mobile networks—a technology overview. IEEE Commun Surv Tutor 17(1):405–426
34. Peng M, Yan S, Zhang K, Wang C (2016) Fog-computing-based radio access networks: issues and challenges. IEEE Netw 30(4):46–53
35. Li D (2015) Opportunistic DF–AF selection for cognitive relay networks. IEEE Trans Veh Technol 65(4):2790–2796
36. Liu Y, Pan G, Zhang H, Song M (2016) Hybrid decode-forward and amplify-forward relaying with non-orthogonal multiple access. IEEE Access 4:4912–4921
37. Zhang Z, Ma Z, Xiao M, Ding Z, Fan P (2016) Full-duplex device-to-device-aided cooperative nonorthogonal multiple access. IEEE Trans Veh Technol 66(5):4467–4471
38. Swami P, Mishra MK, Trivedi A (2016) Performance analysis of two-tier cellular network using power control and cooperation. In: IEEE international conference on advances in computing, communications and informatics (ICACCI), pp 322–327
39. Swami P, Mishra MK, Trivedi A (2017) Analysis of downlink power control and cooperation scheme for two-tier heterogeneous cellular network. Int J Commun Syst. https://doi.org/10.1002/dac.3282
40. Varshney LR (2008) Transporting information and energy simultaneously. In: IEEE international symposium on information theory. IEEE, pp 1612–1616
41. Ding Z, Peng M, Poor HV (2015) Cooperative non-orthogonal multiple access in 5G systems. IEEE Commun Lett 19(8):1462–1465
42. Figueiredo M, Alves LN, Ribeiro C (2017) Lighting the wireless world: the promise and challenges of visible light communication. IEEE Consum Electron Mag 6(4):28–37

43. Yin L, Popoola WO, Wu X, Haas H (2016) Performance evaluation of non-orthogonal multiple access in visible light communication. IEEE Trans Commun 64(12):5162–5175
44. Inan B, Lee SJ, Randel S, Neokosmidis I, Koonen AM, Walewski JW (2009) Impact of LED nonlinearity on discrete multitone modulation. J Opt Commun Netw 1(5):439–451
45. Marshoud H, Kapinas VM, Karagiannidis GK, Muhaidat S (2015) Non-orthogonal multiple access for visible light communications. IEEE Photon Technol Lett 28(1):51–54
46. Yu Y, Giannakis GB (2005) Sicta: a 0.693 contention tree algorithm using successive interference cancellation. In: 24th annual joint conference of the IEEE computer and communications societies, vol 3, pp 1908–1916
47. Yu Y, Giannakis GB (2007) High-throughput random access using successive interference cancellation in a tree algorithm. IEEE Trans Inf Theory 53(12):4628–4639
48. Lee T (1977) The nonlinearity of double-heterostructure LED's for optical communications. Proc IEEE 65(9):1408–1410
49. Mitra R, Bhatia V (2017) Precoded chebyshev-NLMS-based pre-distorter for nonlinear LED compensation in NOMA-VLC. IEEE Trans Commun 65(11):4845–4856
50. Larsson EG, Edfors O, Tufvesson F, Marzetta TL (2014) Massive MIMO for next generation wireless systems. IEEE Commun Mag 52(2):186–195
51. Liu G, Hou X, Jin J, Wang F, Wang Q, Hao Y, Huang Y, Wang X, Xiao X, Deng A (2017) 3-D-MIMO with massive antennas paves the way to 5G enhanced mobile broadband: from system design to field trials. IEEE J Sel Areas Commun 35(6):1222–1233
52. Marzetta TL (2010) Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans Wirel Commun 9(11):3590–3600
53. Ding Z, Adachi F, Poor HV (2015) The application of MIMO to non-orthogonal multiple access. IEEE Trans Wirel Commun 15(1):537–552
54. Ding Z, Schober R, Poor HV (2016) A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment. IEEE Trans Wirel Commun 15(6):4438–4454
55. Senel K, Cheng HV, Björnson E, Larsson EG (2019) What role can NOMA play in massive MIMO? IEEE J Sel Top Signal Process 13(3):597–611
56. Khoshkholgh MG, Navaie K, Yanikomeroglu H (2010) Access strategies for spectrum sharing in fading environment: overlay, underlay, and mixed. IEEE Trans Mob Comput 9(12):1780–1793
57. Song Z, Wang X, Liu Y, Zhang Z (2019) Joint spectrum resource allocation in NOMA-based cognitive radio network with SWIPT. IEEE Access 7:89594–89603
58. Xu W, Qiu R, Jiang X-Q (2019) Resource allocation in heterogeneous cognitive radio network with non-orthogonal multiple access. IEEE Access 7:57488–57499
59. Bastug E, Bennis M, Médard M, Debbah M (2017) Toward interconnected virtual reality: opportunities, challenges, and enablers. IEEE Commun Mag 55(6):110–117
60. Wang F, Xu J, Ding Z (2017) Optimized multiuser computation offloading with multi-antenna NOMA. In: IEEE Globecom Workshops, pp 1–7
61. Kiani A, Ansari N (2018) Edge computing aware NOMA for 5G networks. IEEE Internet Things J 5(2):1299–1306
62. Frew EW, Brown TX (2008) Airborne communication networks for small unmanned aircraft systems. In: Proceedings of the IEEE, vol 96, no 12
63. Mozaffari M, Saad W, Bennis M, Debbah M (2017) Wireless communication using unmanned aerial vehicles (UAVs): optimal transport theory for hover time optimization. IEEE Trans Wirel Commun 16(12):8052–8066
64. Liu Y, Qin Z, Cai Y, Gao Y, Li GY, Nallanathan A (2019) Uav communications based on non-orthogonal multiple access. IEEE Wirel Commun 26(1):52–57
65. Sabharwal A, Schniter P, Guo D, Bliss DW, Rangarajan S, Wichman R (2014) In-band full-duplex wireless: challenges and opportunities. IEEE J Sel Areas Commun 32(9):1637–1652
66. Sun Y, Ng DWK, Ding Z, Schober R (2017) Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems. IEEE Trans Commun 65(3):1077–1091
67. Zhong C, Zhang Z (2016) Non-orthogonal multiple access with cooperative full-duplex relaying. IEEE Commun Lett 20(12):2478–2481
68. Sharma S, Deka K, Bhatia V, Gupta A (2019) Joint power-domain and SCMA-based NOMA system for downlink in 5G and beyond. IEEE Commun Lett 23(6):971–974

Vimal Bhatia (M’99-SM’12) received Ph.D. from the Institute for Digital Communications, The University of Edinburgh, U.K., in 2005. During the work for the Ph.D. degree he received the IEE fellowship for collaborative research on OFDM with Prof. Falconer at Carleton University,Canada. He has been with various telecommunication and new media companies in the U.K. and India. He is currently Professor with IIT Indore, India. He is the author or co-author of around 220 peer reviewed journals and conferences. He also holds eleven published patents. As co-author, he has received the best paper awards at OSA and IEEE conferences, and was awarded the Young Faculty Research Fellow Award from the Ministry of Electronics and Information Technology, India. His research interests are in 5G networks, visible light communication, NOMA, and statistical signal processing. He is serving as an Editor of the IETE Technical Review and is reviewer for IEEE, OSA, Elsevier, and Springer. He is also a certified SCRUM Master.

Pragya Swami completed B.Tech from the Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India, in 2014, and M.Tech from the Indian Institute of Information Technology and Management Gwalior, India, in 2016. Currently, she is pursuing Ph.D. from the Indian Institute of Technology Indore, Indore, India. Her research interests include performance analysis of non-orthogonal multiple access for 5G and beyond network, millimeter-wave communication, heterogeneous and hybrid networks.

Sanjeev Sharma received M.Tech. from the Indian Institute of Technology Guwahati, Guwahati, India, in 2010 and the Ph.D. from the Indian Institute of Technology Indore, Indore, India, in 2018. He is currently an Assistant Professor in the Department of Electronics Engineering at IIT (BHU) Varanasi, India. His main research interests include robust receiver design and sparsity-based signal processing for next-generation wireless communications.

Rangeet Mitra received B.Tech. from Asansol Engineering College, WBUT, India, in 2008, the M.Tech. from Indian Institute of Technology Guwahati, India, in 2010, and the Ph.D. from Indian Institute of Technology Indore, India, in 2017. His research interests include adaptive signal processing in RKHS and nonlinear signal processing for upcoming 5G-based communication systems.