IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 10, 2021, accepted January 14, 2022, date of publication January 25, 2022, date of current version February 1, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3146349

On the Comparison of Optimal NOMA and OMA in a Paradigm Shift of Emerging Technologies

JOYDEV GHOSH[1], (Member, IEEE), IN-HO RA[2], (Member, IEEE), SAURABH SINGH[3], HÜSEYIN HACI[4], KHALED A. AL-UTAIBI[5], (Member, IEEE), AND SADIQ M. SAIT[6], (Senior Member, IEEE)

[1]School of Computer Science and Robotics, Tomsk Polytechnic University, 634050 Tomsk, Russia
[2]School of Computer, Information and Communication Engineering, Kunsan National University, Gunsan 54150, South Korea
[3]Department of Industrial and Systems Engineering, Dongguk University, Seoul 04620, South Korea
[4]Department of Electrical and Electronics Engineering, Near East University, 99138 Mersin, Turkey
[5]Computer Science and Software Engineering Department, University of Ha'il, Ha'il 53962, Saudi Arabia
[6]Department of Computer Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

Corresponding author: In-Ho Ra (ihra@kunsan.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government [Ministry of Science and ICT (MSIT)], under Grant 2021R1A2C2014333.

ABSTRACT Non-orthogonal multiple access (NOMA) is a better multiple access technique than orthogonal multiple access (OMA), precisely orthogonal frequency division multiple access (OFDMA) scheme, at the conceptual level for fifth-generation (5G) networks and beyond 5G (B5G) networks. We investigate the potentials of the schemes by comparing the proposed NOMA scheme with the traditional cooperative communication NOMA (CCNOMA) scheme, rather than the comparison between NOMA and OMA only. To probe the effectiveness of NOMA as a multiple access technique, we propose a novel NOMA scheme considering two adjacent BSs with a special design of the transceiver architecture. The proposed scheme provides a reasonable data rate to both near user (NU) and far user (FU) without compromising the quality of service (QoS) to anyone of them. The conclusive analyses on the optimization framework of multi-user sum rate, capacity, transmit power, spectral efficiency (SE), and energy efficiency (EE) trade-off for NOMA and OFDMA schemes have been established to a succession of derivations. Under the analytical optimization framework, we also prove quite a few properties for them. Simulation results confirm the theoretical findings and show that the two schemes can efficiently approach the optimal power allocation, minimization of power consumption, and optimal SE-EE trade-off, and the proposed NOMA scheme provides comparatively better data sum rates than the baseline OMA scheme.

INDEX TERMS Non-orthogonal multiple access (NOMA), orthogonal multiple access (OMA), orthogonal frequency division multiple access (OFDMA), multi-user sum rate, capacity, transmit power, spectral efficiency (SE), energy efficiency (EE), multi-objective weighted sum optimization.

I. INTRODUCTION

Due to today's high volume video services and massive annual growth of the Internet of things (IoT) devices, International Telecommunication Union (ITU) for International Mobile Telecommunications (IMT)-2020 and beyond is set to fulfill these diverse requirements by 5G and beyond. The orthogonal multiple access (OMA) scheme is a reasonable choice to match up with the expected performance requirement for the first to the fourth generation (1G to 4G) wireless networks. OMA as multiple access (MA) scheme is well exploited in 4G long term evolution (LTE) and LTE-Advanced (LTEA) networks [1], [2]. As the consideration of ultra-dense network (UDN) configurations, millimeter wave (mmWave), and massive multiple-input multiple-output (MIMO) inadequate, thus one of the keys to achieving the success to the future radio access (FRA) in the 2020s is to be non-orthogonal multiple access (NOMA) scheme. Many of the personnel from industry and academia sprang up to exploit this advanced MA technology.

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaofan He.

A. MOTIVATION AND RELATED WORK

Unlike OMA that allows each user to access the allocated time-frequency resources exclusively, NOMA allows multiple users to simultaneously use the same time-frequency resources with different power levels. The throughput and the delay of Buffer-aided cooperative NOMA were discussed in [3], whereas the sensing accuracy under machine learning-enabled solutions was probed to deal with cooperative spectrum sensing (CSS) issue for NOMA [4]. New wireless powered Internet of Things (IoT) network design based on the principle of NOMA plus time division multiple access (TDMA) can further maximize the sum-throughput [5]. Joint operation to maximize spectral efficiency (SE) is also possible between Orthogonal frequency division multiplexing (OFDM) and NOMA, but the OFDM-NOMA scheme has a serious setback of a high peak-to-average power ratio (PAPR) [6]. The application of cooperative NOMA to the far user causes extra power dissipation, cooperative simultaneous wireless information, and power transfer (SWIPT) NOMA protocol are proposed in [7], [8] to compensate power dissipation by using the energy harvesting (EV) technique. To achieve high SE, random access with layered preambles (RALP) based on the concept of NOMA to support different types of devices has been proposed in [9]. [10] studied the asymptotic ergodic rates of two users for full-duplex (FD)/half-duplex (HD) NOMA with its newly derived closed-form expressions. In [11], the upper and lower bound of ergodic sum-rate under Nakagami-m fading channels were analyzed in NOMA for amplify-and-forward (AF) relaying networks. [12] carried out the analytical investigation of NOMA systems based on imperfect channel state information (CSI) and the known second-order statistics (SoS) of the channel. In [13], the authors proposed cache-aided NOMA mobile edge computing (MEC) framework and the application of reinforcement learning (RL). The MEC-based IoT network allows the devices to release the excessive burden of computation offloading on the MEC server and it takes less time to finish the task which in turn certainly helps to reduce energy consumption [14], [15]. Ding *et al.* in [16] shown that NOMA-MEC can achieve significant energy gain compared to OMA-MEC under low-latency constraints. In [17], the authors proposed the design of NOMA-MEC in presence of the external eavesdropper subject to the secrecy offloading rate constraints. [18] applied NOMA on mmWave multicasting for the Poisson point process (PPP) and derived the theoretical closed-form expressions of sum-rates, while [19] characterizes the performance of two users NOMA system model by considering the stochastic geometry technique. [20] investigated Physical-layer security (PLS) of large-scale networks for secure NOMA transmission with artificial noise. [21] demonstrated the analytical findings of feedback system for both user ordering and radio access technology (RAT) rather than for unlike user ordering only.

Unlike traditional OMA, NOMA can achieve higher SE due to its capability to exploit the channel diversity more effectively via SIC [22]. With the help of achievable sum-rate regions' comparison between NOMA and OMA, the authors in [23] proved the primacy of NOMA over OMA and it is also valid for randomly deployed users' scenario [24]. It had been shown in [25] that MIMO-NOMA for a two-user multi-cluster holds a larger sum-rate advantage over MIMO-OMA, while [26] extended its validation for a multi-user per cluster scenario; in fact, this is also conclusive for mmWave massive MIMO system [27], wireless networks with no energy harvesting [24] and SWIPT networks [28]. The authors studied a multi-carrier NOMA and a single-carrier NOMA with suitable sub-channel allocation and power allocation in [29], [30] for energy-efficient transmission design. [31] introduced a novel transmission scheme by synthesizing blind interference alignment (BIA) and NOMA, namely B-NOMA, and simulation results therein confirm that the B-NOMA scheme for the multi-user sum-rate performs better than regularized zero-forcing (RZF) based MISO-OMA and MISO-NOMA schemes.

In [32], the authors tried to deal with the different design challenges on channel acquisition, the realization of precoder at the transmitter end and SIC at the receiver end by introducing a novel practical DL-NOMA scheme for wireless local area networks (WLANs) on a wireless testbed platform. [33] discussed a novel NOMA-based Multi-way relay networks (MWRNs) protocol and derived its closed-form expressions for the multi-user sum rate and the EE. There are many challenges in the areas of imperfect synchronization, interference mitigation, power allocation strategy, maximization of sum-rates, SE, and EE, that are yet to be addressed for wireless backhaul [34]–[37].

With the new degree of freedom, intelligent reflecting surfaces (IRSs) aided NOMA design can be a cost-effective solution to increase the SE and EE of B5G networks [38]. In [39], [40], the IRS was brought into effective action to achieve the capacity region with high possibility and also to maximize communication coverage and EE by assisting a far user (FU) in secure data transmission strategy, where this FU is paired with a near user (NU) via NOMA scheme. Nowadays, ambient backscatter communication (AmBC) has drawn tremendous attention due to its ability of radio RF energy harvesting (EH) and extreme low power consumption which makes it a potential solution to green Internet-of-Things (IoT) networks [41]. In [42], authors proposed a novel backscatter cooperation (BC)-NOMA scheme and then compared it with non-cooperation (NC)-NOMA, conventional relaying (CR)-NOMA, and incremental relaying (IR)-NOMA with the conclusion that the proposed scheme strictly outperformed other schemes. More precisely, by tuning the load impedance, NU in the BC-NOMA scheme divides a fraction of its received signal for original signal reconstruction and backscatters the rest fragment to enhance the signal to interference plus noise ratio (SINR) at FU.

To achieve high reliability in Ultra-Reliable Low Latency Communications (URLLC), it is highly desirable to set a target block error rate (BLER) for B5G to $[10^{-10} - 10^{-6}]$, whereas the requirement of this typical value to be 10^{-3} for 4G [43]. Thus a B5G network has to rely on hybrid automatic repeat request (HARQ) retransmissions and channel coding (e.g., Turbo code). [44] shown that NOMA with re-transmission via HARQ is a better choice compared to that of OMA with HARQ. NOMA-assisted fog-radio access networks (F-RAN) in [45]–[47] can be a very useful approach in the design of high-performance communication systems, specifically for URLLC traffic and enhanced Mobile Broadband (eMBB), with the help of energy-efficient resource allocation.

B. CONTRIBUTIONS AND ORGANIZATION

From the above motivations, the contributions of this paper are summarized as follows:

- At first, we study two conventional schemes such as the OFDMA scheme apart from OMA and cooperative communication NOMA (CCNOMA) scheme with their pros and cons. Some of the salient features, such as massive connectivity, low latency, and signal processing cost, and traffic volume in NOMA have been discussed. Thereafter, we proposed a novel NOMA system model with a special transceiver design. The proposed NOMA scheme takes into account both power domain and interference cancellation combining (ICC) despite traditional successive interference cancellation (SIC) only. There may be many advantages and disadvantages of all those schemes, but the prime objective of this work is successfully done by comparing three schemes in a generalized manner with problem formulation from the optimization viewpoint for optimal power allocation, minimization of circuit power consumption, data rates, capacity, and SE-EE trade-off.
- During the multi-user sum-rate optimization, it has been proved that the proposed NOMA scheme provides much better data rates for two users compared to the OFDMA scheme and achievable data rate for user1 is strictly concave function regardless of any schemes. We then derive the closed-form expressions of the capacity for the proposed NOMA scheme and traditional OFDMA scheme, respectively. Moreover, regarding the problem formulation of transmit power optimization with a given outage probability (OP) constraint, we first derive optimal SIC decoding order to provide feasibility conditions. The multi-objective optimization problem for SE and EE of the schemes has been formulated and able to conclude that the EE function is a strictly quasi-concave concerning the transmit power. We extend optimization problem formulation for SE and EE employing the weighted sum method to further shed light on their trade-off and also derive its closed-form expression.
- For both NOMA and OMA, we present the trade-off between the data rate of user 1 and user 2, and EE and SE, respectively, to highlight the quantifiable characteristics of the NOMA principle being a strong candidate to be employed at the fifth generation (5G) and beyond the network. We present a performance comparison of NOMA in HetNets, considering OMA as a benchmark. The results indicate that up to around 25% increase in capacity can be obtained by applying NOMA despite OMA. We further show that the data rate of user 1, SE and EE increase, whereas the data rate of user 2 decreases, whilst the working principle moves toward NOMA from OMA. Besides, the trade-off results proactively explore that NOMA performs better in contrast to OMA in the upper SE regime and upper user1's data-rate regime, respectively, irrespective of channel condition.

II. ORTHOGONAL MULTIPLE ACCESS (OMA)

Based on IEEE 802.16 standard, orthogonal multiple access schemes - Orthogonal frequency division multiple access (OFDMA) and time-division multiple access (TDMA) - can be adapted for the mid-band spectrum, especially bands below 5 GHz which will play an integral role for 5G [52]. OFDMA is preferable of the two OMA schemes for deployment since it can achieve flat-fading, narrow-band subcarriers and utilize the frequency spectrum more effectively through overlapping sub-carriers.

A. ORTHOGONAL FREQUENCY DIVISION MULTIPLE ACCESS (OFDMA)

The assumed system model of OFDMA is developed with n_{T_x} and n_{R_x} transmit and receive antennas respectively and allowed to transfer $1 \leq n_{seq} \leq \min(n_{T_x}, n_{R_x})$ sequences of data bits. The mapping of data column vector $x \in \mathbb{C}^{n_{seq} \times 1}$ at T_x side is done with the consideration of n_{T_x} transmit antennas and $F \in \mathbb{C}^{n_{T_x} times n_{seq}}$ precoding matrix. We further use subscript w to denote wanted signal and u to denote unwanted signal as the interference at R_x side. The data column vector at R_x with $G \in \mathbb{C}^{n_{R_x} \times n_{seq}}$ combining matrix is given by,

$$v = \acute{h}_w x_w + \acute{h}_u x_u + N_o, \tag{1}$$

where $N_o \in \mathbb{C}^{n_{R_x} \times 1} \sim \mathcal{N}(0, \sigma^2 I_{n_{R_x}})$ denotes additive white gaussian noise (AWGN) at the receiver, which is independent and identically distributed (i.i.d.), and has a zero-mean with σ^2 variance in which σ stands for standard deviation and $I_{n_{R_x}} \in \mathbb{C}^{n_{T_x} \times n_{R_x}}$ indicates the identity matrix. $x_u = [x_{u_1}^T, x_{u_2}^T, \cdots, x_{u_k}^T]^T$ and $\acute{h}_u = G_u F_u H_u = [\acute{h}_{u_1}, \acute{h}_{u_2}, \cdots, \acute{h}_{u_k}]$ denote the string of the k number of unwanted interference signals at R_x side and their associated channels respectively, where $H_u \in \mathbb{C}^{n_{T_x} \times n_{R_x}}$ denotes respective fading channel matrix and $(.)^T$ represents the transpose operator. Similarly, $\acute{h}_w = G_w F_w H_w$ denotes estimation of the equivalent channel matrix for the wanted signal, where $H_w \in \mathbb{C}^{n_{T_x} \times n_{R_x}}$ denotes respective fading channel matrix.

The Low mean squared deviation (LMSD) combining matrix can be expressed as below in the context of the

estimation for x_w,

$$g = (\acute{h}_w \acute{h}_w^H + V)^{-1} \acute{h}_w, \tag{2}$$

where $(.)^T$ represents the hermitian transpose operator and V denotes receiver design metric.

The estimation of x_w can be expressed by,

$$\acute{x}_w = g^H v. \tag{3}$$

Case Study I: For interference cancellation combining (ICC) receiver, the receiver design metric can be expressed by,

$$V = \mathbb{E}\left[\acute{h}_u \acute{h}_u^H\right] + \sigma^2 I_{n_{R_x}}. \tag{4}$$

The design metric of the ICC receiver is achieved by estimating the covariance matrix of the unwanted interference signals. The application of ICC-receiver is specifically advantageous for the small cell deployed uncoordinated high-dense networks due to its potential in reducing the mean squared deviation (MSD) significantly.

Case Study II: For maximal ratio combining (MRC) receiver, the receiver design metric can be expressed by,

$$V = \mathbb{E}\left[\sum_{i=1}^{k}\sum_{j=1}^{n_{seq}} |\acute{h}_{u_{i,j}}|^2\right] + \sigma^2, \tag{5}$$

The design metric of the MRC-receiver is estimated by evaluating the complete unwanted power associated with interference signal and noise at each antenna apart from the receiver side.

The signal to interference plus noise ratio (SINR) for the j^{th} sequence of data bits can be expressed by

$$\gamma_j = \frac{g_{w(j)} g_{w(j)}^H H_{w(j)} H_{w(j)}^H}{g_{w(j)} \left(\acute{h}_{u(j)} \acute{h}_{u(j)}^H + H_{u(j)} H_{u(j)}^H\right) g_{w(j)}^H}. \tag{6}$$

The SINR can then be normalized and expressed as below in order to design imperfect transceiver,

$$\gamma_{norm} = \frac{\gamma_j \gamma_{max}}{\gamma_j + \gamma_{max}}. \tag{7}$$

The OFDMA scheme follows the orthogonal user multiplexing, where the bandwidth of $\beta(0 < \beta < 1)$ is allocated to the near user (NU) and the rest $1 - \beta$ is allocated to the far user (FU). Therefore, data rates of two UEs can be expressed as below,

$$R_1^{OFDMA} = \beta \log_2\left(1 + \frac{\delta \gamma_{norm} |Ph_1|^2}{\beta \sigma_1^2} \lambda_1\right), \tag{8}$$

$$R_2^{OFDMA} = (1-\beta) \log_2\left(1 + \frac{(1-\delta)\gamma_{norm} |Ph_2|^2}{(1-\beta)\sigma_2^2} \lambda_2\right), \tag{9}$$

where λ_1 and λ_3 denote the respective eigenvalues of the channel matrices for NU and FU.

III. NON-ORTHOGONAL MULTIPLE ACCESS (NOMA)

To date, high SE, EE and user data-rate, and massive connectivity requirement of 5G wireless communications can not be fulfilled by adopting OMA discussed in Section II. Different from conventional OMA, NOMA can provide better network performance with adequate non-orthogonal resource allocation. Different from other surveys of NOMA, this section focuses on principles of customized NOMA, salient features, and pros/cons of NOMA as follows;

Massive Connectivity: Many UEs are allowed to share the same resource block simultaneously in the power domain of the NOMA scheme at the cost of inter-UE interference in each resource block, however in case the inference is managed well this helps to improve SE. To deal with such challenging multi-user detection (MUD), SIC scheme have been developed to apply at the receiver side to aid decoding the superimposed signals. Among various NOMA techniques, there are schemes introduced with comparatively lower receiver complexity and significantly improved performance which makes it suitable for providing massive connectivity for internet of things (IoT) applications. The main advantage of NOMA for this concern is that at OMA the maximum number of users communicating simultaneously at a scheduling time is limited to the number of sub-carriers. However at NOMA, due to frequency reuse within a cell, this number can be much larger than the number of sub-carriers. This can be an enabling technology to achieve massive simultaneous connectivity.

Low Latency and Data Overhead: In multi-carrier OMA transmission, all the active users compete for sub-carrier allocation through the scheduling mechanism. Since a one-to-one allocation is done between users and sub-carriers, massive connectivity greatly increases the competition and latency and becomes infeasible with the limited number of sub-carriers. With NOMA, many-to-one allocation can be achieved and competition and latency can be considerably reduced. Further, since each transmitted symbol can carry multiple users data with a single header, the data overhead can be minimized.

The combination of NOMA and mobile edge computing (MEC) can be potential candidates in today's networks and the model of their joint system for two users' case studies can certainly help to provide low latency and high EE in MEC offloading. As far as the practical scenario is concerned, more than two users or even massive users case studies are well suited. During the operation of massive users with only one MEC server, NOMA offers multiple users to finish the offloading task simultaneously and ensures low offloading latency [48].

Traffic Offloading: With the usage of the vast amount of digital data due to video applications these days, the data traffic volume is growing dramatically. Offloading traffic from macro-cells in a seamless and efficient way become a vital point for future wireless networks. To fulfill such demands, 5G compared to 4G is expected to come up with the higher potential to deal with traffic volume along with much faster data speed (up to 2 Gbps), low latency (about 3-10 ms),

and better stability. Hence, data traffic volume offloading is becoming an important design metric in a paradigm shift of emerging networks. With the assumption that the average estimated data load of the macro-cells is 65% at the peak hours and if the peak hour can bear 25% of daily traffic, the data traffic volume can be defined as follows,

$$V_{tv} = \frac{\left(\frac{R\times 3600\times 65}{25}\right)\times 365}{\mathrm{A}}\ [Gb/\mathrm{Year}/\mathrm{Km}^2],$$

where R denotes data rate capacity in Gbps, A denotes area in square kilometer. The enourmous amount of traffic volume at a macro-cell can be calculated from this formula.

In a massive connectivity scenario, at peak hours the system may easily be near-congestion or congested stated very often. Thus, offloading to small-cells becomes a crucial aspect for future wireless communications. This paper discusses and provides performance comparison for small-cell inspired NOMA, cooperative-communication NOMA and a proposed NOMA scheme to elaborate on this vital aspect.

A. SMALL CELL INSPIRED NOMA

The OMA schemes have been applied for interference mitigation in 4G networks, for instance, OFDMA for downlink (DL) transmission and SC-FDMA for uplink (UL) transmission. Nowadays, Small Cell enabled 5G networks are anticipated to accomplish keeping large SE, large EE, ultrahigh connectivity, less latency, and unusual handoffs. To fix these issues, the NOMA scheme has been probed deliberately and positively to become a probable substitute to OFDMA and SC-FDMA for Small Cell enabled 5G networks.

A main feature of the NOMA technique is to facilitate more than one portable UE over the same frequency spectrum applying non-orthogonal resource allocation. As NOMA brings about a manageable quantity of UE interference, this can be reduced with the help of advanced UE demodulators at the price of an even more complex receiver blueprint. Due to the numerous merits of the NOMA techniques, they have been explored as promising radio access candidates in 5G. There are various types of NOMA schemes, such as power-domain multiplexing, code-domain multiplexing, pattern division multiple access (PDMA), bit division multiplexing (BDM). Besides, this paper also investigates how NOMA functions while this is integrated with other wireless technologies, for instance, cooperative communications, MIMO, and beamforming. A probable DL NOMA-enabled 5G networks composes of many small cell base stations (SCBSs) installed along the roadsides, where a set of mobile objects moving on the road. All SCBSs is linked to the server fixing by the wireless service provider through fiber. The introduced network makes use of NOMA with SIC as a radio access technique. In NOMA with SIC, each SCBS chooses to take up concurrent transmission for more than one user applying SC and each mobile device chooses to take up the SIC receiver to decode the intelligence from its allied SCBS. Nowadays the greater degree of Small Cell densification is becoming the ultimate option for 5G networks. At the same time, the experience of frequent handoff execution invokes the cell partnership optimization scheme can efficiently decrease the handoff rate [52]. Since handoffs set off a total host of complex mechanisms, which leads to unwelcome handoff delays. This delay may appreciably deteriorate the transmission quality of performing consistently well for mobile objects. Thus, describing the distinctive nature or features of the handoff delay should be a high priority to design handoff schemes in partnership with NOMA-enabled Small Cell Networks (SCNs). The subject of an optimization plan of action, and outlook in [53] introduces the mobility-aware cell partnership. To decrease the handoff complexity, duplex connectivity that permits the mobile user to be concurrently allied with Macrocell Base Stations (MBSs) and SCBSs is systematized as a way out for aiding the high-speed UE mobility in SCNs. Duplex connectivity authorizes UEs to carry on the link to the MBS and therefore assurances the durable performance because of the broad network coverage of MBSs, however, UEs may hand off with other SCBSs often.

B. COOPERATIVE COMMUNICATION NOMA (CCNOMA)

The main feature of NOMA with cooperative communication is that UE with a powerful channel state performs as a relay to help out UEs with infirm channel state. Furthermore, two UEs DL scenarios have been shown in Figure 1 as an instance. An operating mechanism of NOMA-enabled cooperative communication can be classified into two modes, such as direct communication mode and cooperative communication mode, respectively. In the time of direct communication mode, the BS transmits a composite of the piece of information for UE X (denotes infirm channel state) and UE Y (denotes powerful channel state). In the time of cooperative communication mode, after bringing about SIC at UE Y for decoding UE X's information, subsequently UE Y performs as a relay to pass on the decoded intelligence to UE X. Hence double copies of the piece of information are picked up at UE X via different channels. However, an extensive and new NOMA-enabled cooperative communication technique incorporating K UE was presented in [53]. An application of SIC at receiver in NOMA enabled networks puts cooperative communication in a favorable or superior position and result in the information to the UEs with infirm channel states that have previously been decoded by the UEs with powerful channel states. Thus this is quite normal to enlist the UEs with powerful channel states as relays. As a result, the action or process of receiving the UEs with infirm channel states is remarkably upgraded. As presented in [22], NOMA enabled cooperative communication performs better than NOMA enabled non-cooperative communication in respect of the outage probability of the UE pairing and the underprivileged UE, respectively.

C. PROPOSED NOMA SCHEME

In figure 2, two base stations and three user equipment (UE) are considered in the system model to propose the principle of

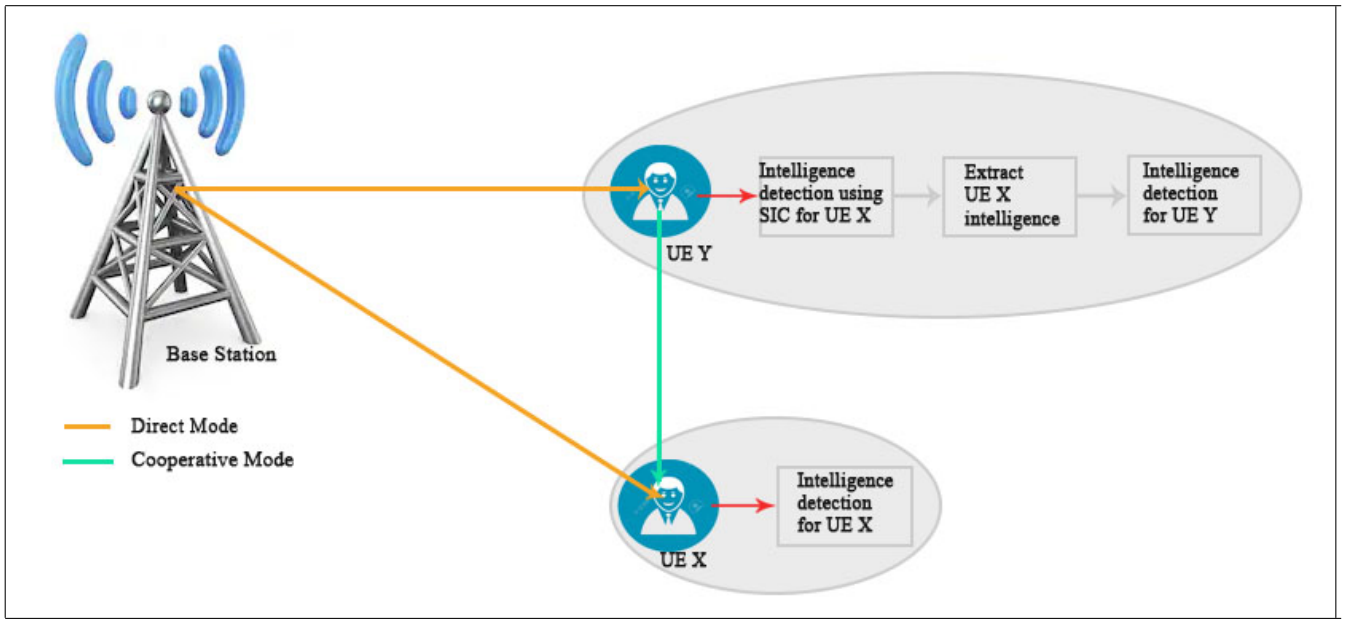


FIGURE 1. SIC operation in direct and cooperative communication modes [22].

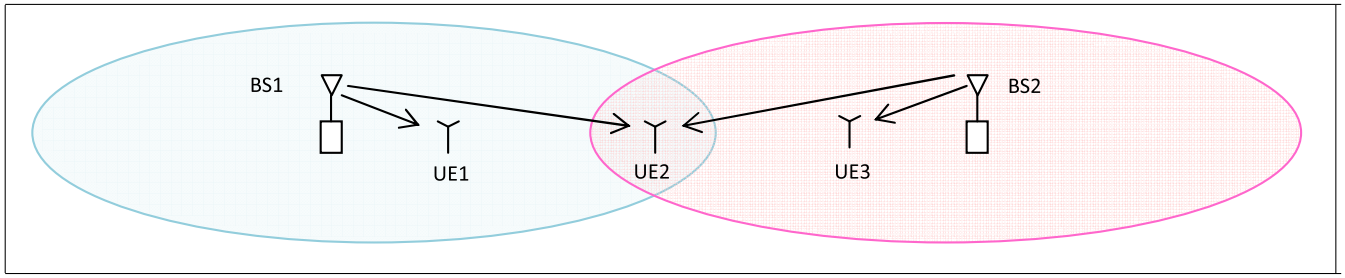


FIGURE 2. System model of proposed NOMA scheme.

the NOMA scheme. In the given scenario, UE1 and UE3 are assumed as near users (NUs) to BS1 and BS2 respectively and UE2 as far user (FU) to both BSs. By following the superposition coding method, both BSs send data symbols to the FU and each of the NUs also receives data symbols from their respective BS respectively. The channel coefficient between a BS j and a UE i is symbolized by $h_{i,j}, \forall i \in \{1, 2, 3\}, \forall j \in \{1, 2\}$. The used notation to denote data symbol is $x_i, \forall i \in \{1, 2, 3\}$ and $\acute{x}_2$ is precisely used to denote data symbol to UE2 from BS2.The received signal at the NUs and the FU are, respectively, given by

$$y_1 = h_{11}(x_1 + x_2) + h_{12}(x_3 + \acute{x}_2) + n_1, \quad (10)$$
$$y_2 = h_{21}(x_1 + x_2) + h_{22}(x_3 + \acute{x}_2) + n_2, \quad (11)$$
$$y_3 = h_{31}(x_1 + x_2) + h_{32}(x_3 + \acute{x}_2) + n_3, \quad (12)$$

where y_i and n_i denote received signal and noise at UE $i, \forall i \in \{1, 2, 3\}$. Assuming $p_i = \mathbb{E}[|x_i|^2], \forall i \in \{1, 2, 3\}$, $p_2 = \mathbb{E}[|x_2|^2] + \mathbb{E}[|\acute{x}_2|^2] = 2\mathbb{E}[|x_2|^2]$ as assigned power is equal to x_2 and $\acute{x}_2$. Hence, p_1 and p_3 are becoming considerably smaller than p_2.

By applying ICC, (1) at UE 1 becomes

$$y_1 - (h_{11}x_2 + h_{12}\acute{x}_2) = h_{11}x_1 + h_{12}x_3 + n_1 \quad (13)$$

As associated power of x_1 and x_3 are very much alike and $h_{11} \gg h_{12}$, hence it is not possible to decode x_3 and $(h_{12}x_3 + n_1)$ will be treated as unwanted signal in the evaluation of data rate of UE1. Therefore, data rates of UE1 and UE3 can be expressed as

$$R_1^{NOMA} = \mathbb{E}\left[\log_2\left(1 + \frac{|h_{11}|^2 p_1}{\mathbb{E}[|h_{12}|^2]p_3 + \sigma_1^2}\lambda_1\right)\right], \quad (14)$$
$$R_3^{NOMA} = \mathbb{E}\left[\log_2\left(1 + \frac{|h_{32}|^2 p_3}{\mathbb{E}[|h_{31}|^2]p_1 + \sigma_3^2}\lambda_3\right)\right], \quad (15)$$

where λ_1 and λ_3 denote the respective eigenvalues of the channel matrices for UE1 and UE3. As x_2 is decodable at UE1 and UE3 for ICC, the data rates of UE2 can be expressed

as follows,

$$\dot{R}_2^{NOMA} = R_2^{NOMA}\Big|_{UE1} = \mathbb{E}\left[\log_2\left(1+\frac{(|h_{21}|^2+|h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3+\sigma_2^2}\lambda_2\right)\right], \quad (16)$$

$$\ddot{R}_2^{NOMA} = R_2^{NOMA}\Big|_{UE3} = \mathbb{E}\left[\log_2\left(1+\frac{(|h_{21}|^2+|h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{21}|^2]p_1+\sigma_2^2}\lambda_2\right)\right]. \quad (17)$$

where λ_2 denotes the respective eigenvalues of the channel matrices for UE2.

Thus, the maximum data rate of UE2 can be upper-bounded by,

$$R_{2(max)}^{NOMA} = \min\{\dot{R}_2^{NOMA}, \ddot{R}_2^{NOMA}\} \quad (18)$$

The far user (FU) data rate is an important concern in practice, but the OMA scheme does not promise this. Hence, a novel NOMA scheme is proposed to address the FU's data rate to maintain quality of service (QoS). In this context, we have considered the fairness index [49] to measure user fairness and it can be expressed as below,

$$J = \frac{\left[\sum_{i=1}^{3} R_i\right]^2}{3\sum_{i=1}^{3} R_i^2} \quad (19)$$

D. PROPOSED SECTION OF NOMA-TRANSCEIVER

The proposed interference cancellation combining (ICC), which consists of spatial filter (SF) and successive interference cancellation (SIC), can be applied at the receiver with the exploitation of the following points,

- SIC is very useful to deal with co-cell UE multiplexing. The interference mitigation between UEs belonging to the same cell or same cluster is performed by assigning the precoding weights of an identical type. In the SIC process, the receiver needs to be decoded the received message transmit to other UEs for interference mitigation. Therefore, the signal to interference plus noise ratio (SINR) of the message to be eliminated must be large to an adequate degree to make the respective message decodable. The optimum decodable order is followed in descending order of normalized channel gain. Any UE can successfully decode the message of other UE by obeying the occurrence of optimum decodable order.

- SF is an effective tool to manage cross-cell UE multiplexing. The interference mitigation between UEs belonging to the different cells or different clusters is performed by assigning the precoding weights of non-identical types. The prime advantage of SF is that the interference message is attenuated directly by combining the received messages from multiple antennas at the receiver regardless of decoding UE and the channel gains.

Although, it has some major drawbacks that its performance gets deteriorated to a certain extent in the scenario where interference has less effect or non-existent. The response becomes progressively worse due to the reason that the direct measurement produces some amount of artificial interferers in the noise covariance matrix and there may be a fraction of the desired message power gets deducted needlessly.

Besides, SF performance can have a remarkable impact on the following main determinants: -Number of receiving antennas and interferers, -Internet of Things (IoT) level, -Time synchronization, -Interferer channel profile, and -Scheduler.

IV. MULTI-USER SUM RATE OPTIMIZATION

For the proposed NOMA scheme, the achievable multi-user sum rate optimization problem for the power vector, $\rho = [p_1, p_2]^T$, can be formulated as follows,

$$\max_{\rho\geq 0} \sum_{i=1}^{2} R_i, \quad \text{Subject to: } \sum\nolimits_{i=1}^{2} p_i \leq P_{max}. \quad (20)$$

Proposition 1: The problem stated in (19) is to be optimized if and only if

$$p_i = 0, \quad \text{for } i = 2, \quad (21)$$

where $i = 2$ denotes FU of proposed NOMA scheme.

Proof: For the proposed NOMA system model, with the consideration of NU (i.e., UE1 and UE3) and FU (i.e., UE2) we can have

$$\sum_i p_i = P, \quad \forall i \in \{1,2\} \text{ or } \forall i \in \{3,2\} \quad (22)$$

where P denotes maximum transmit power by either BSs. Here p_1, and p_3 denote allocated power to the NUs and p_2 denote allocated power to the FU. Hence, two user sum rate for the cell associated with BS1 can be expressed by

$$f(p_1,p_2) = \mathbb{E}\left[\log_2\left(1+\frac{|h_{11}|^2 p_1}{\mathbb{E}[|h_{12}|^2]p_3+\sigma_1^2}\lambda_1\right)\right] + \mathbb{E}\left[\log_2\left(1+\frac{(|h_{21}|^2+|h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3+\sigma_2^2}\lambda_2\right)\right] \quad (23)$$

Likewise, two user sum rate for the cell associated with BS2 can be expressed by

$$f(p_3,p_2) = \mathbb{E}\left[\log_2\left(1+\frac{|h_{32}|^2 p_3}{\mathbb{E}[|h_{31}|^2]p_1+\sigma_3^2}\lambda_3\right)\right] + \mathbb{E}\left[\log_2\left(1+\frac{(|h_{21}|^2+|h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{21}|^2]p_1+\sigma_2^2}\lambda_2\right)\right] \quad (24)$$

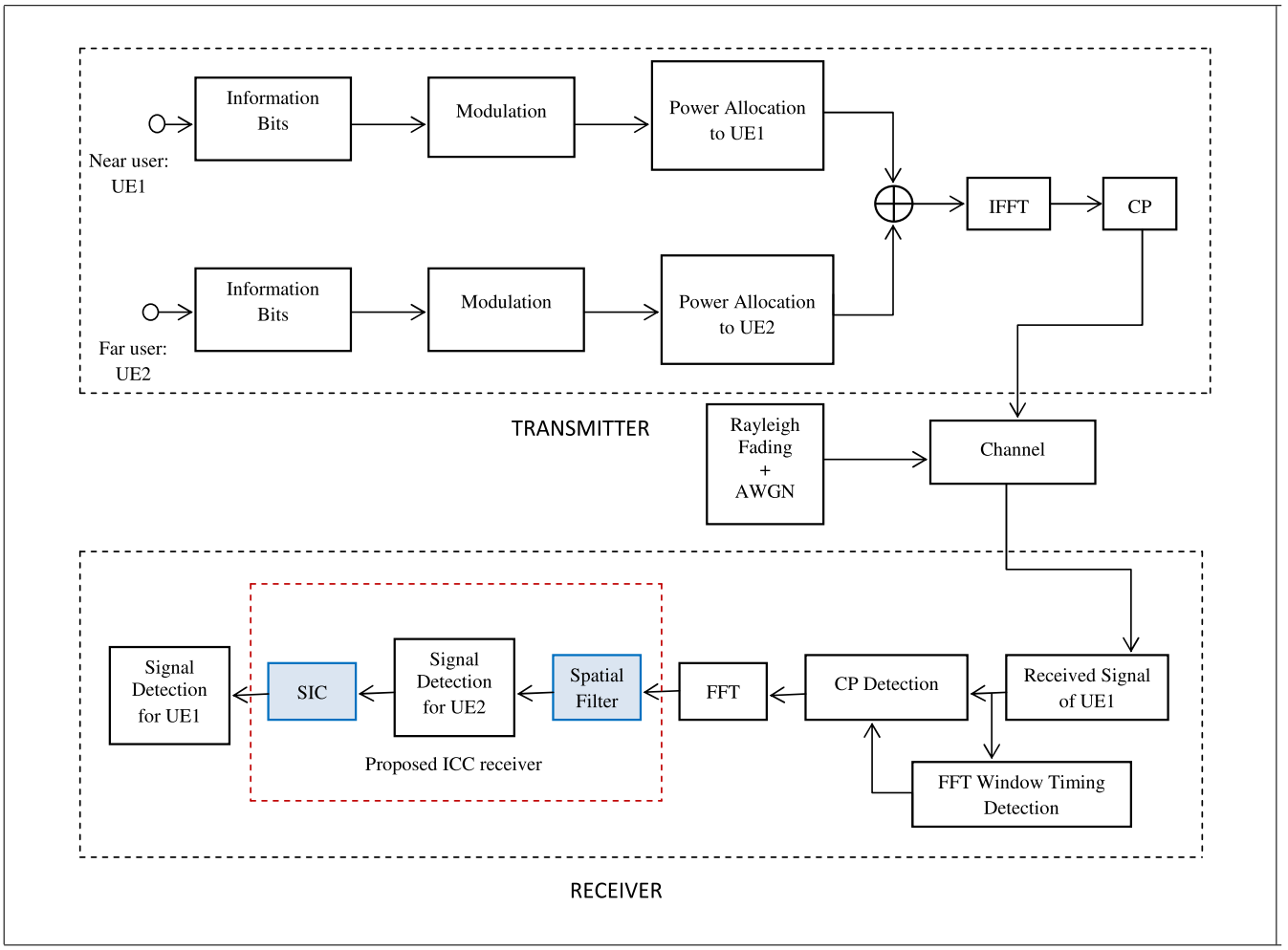


FIGURE 3. Proposed transceiver architecture.

The derivative of (24) with respect to p_1 can be written as[1] in (25).

$$\frac{df(p_1,p_2)}{dp_1} = \frac{1}{\log_e 2}\Bigg[\frac{(|h_{12}|^2p_1+\sigma_1^2)|h_{11}|^2\lambda_1-|h_{11}|^2p_1\lambda_1|h_{12}|^2}{(|h_{12}|^2p_1+|h_{11}|^2p_1\lambda_1+\sigma_1^2)(|h_{12}|^2p_1+\sigma_1^2)} + \frac{2(|h_{22}|^2p_1+\sigma_2^2)\{-(|h_{21}|^2+|h_{22}|^2)\lambda_2\}-(|h_{21}|^2}{\{2(|h_{22}|^2p_1+\sigma_2^2)+(|h_{21}|^2+|h_{22}|^2)(P-p_1)\}} \frac{|h_{22}|^2)(P-p_1)\lambda_2|h_{22}|^2}{\{2(|h_{22}|^2p_1+\sigma_2^2)\}}\Bigg],$$

for $p_3 \approx p_1$ and $p_2 = P - p_1$

$$= \frac{1}{\log_e 2}\Bigg[\frac{|h_{11}|^2\lambda_1\sigma_1^2}{(|h_{12}|^2p_1+|h_{11}|^2p_1\lambda_1+\sigma_1^2)(|h_{12}|^2p_1+\sigma_1^2)} - \frac{(|h_{21}|^2+|h_{22}|^2)\lambda_2}{2(|h_{22}|^2p_1+\sigma_2^2)}\Bigg]\Bigg|_{p_1=P} \tag{25}$$

Therefore, $\frac{df(p_1,p_2)}{dp_1}$ is positive at all times.

$$\text{Thus, } \max_{p_1,p_2\geq 0} f(p_1,p_2) = f(P,0), \tag{26}$$

which confirms the proof that the two-user sum rate of the proposed NOMA scheme is optimized by assigning P_{max} power to the NU, hence the power vector is becoming $\rho = [P_{max}, 0]^T$. ■

Proposition 2: If $p_3 \approx p_1$, data rate of UE1 achieved by both NOMA and OFDMA schemes, respectively, is becoming strictly concave function. But the data rate of UE1 and UE2 for the NOMA scheme is always higher compared to the OFDMA scheme.

Proof: We use the notations (R_1^{NOMA}, $\dot{R}_2^{NOMA}$) and (R_1^{OFDMA}, R_2^{OFDMA}) of UE1 and UE2 for NOMA and OFDMA schemes, respectively. The two user data rates have obtained by accessing the full-resources, i.e., $p_1 + p_2 = P$ for the proposed NOMA scheme and $\beta_1 + \beta_2 = 1$ for the proposed OFDMA scheme. Thus the derivative of (14) with respect to p_1 assuming $p_3 \approx p_1$ can be written as,

$$\frac{dR_1^{NOMA}}{dp_1} = \frac{1}{\log_e 2}\Bigg[\frac{|h_{11}|^2\lambda_1\sigma_1^2}{(|h_{12}|^2p_1+|h_{11}|^2p_1\lambda_1+\sigma_1^2)(|h_{12}|^2p_1+\sigma_1^2)}\Bigg] \tag{27}$$

[1] $\log_a m = \frac{\log_b m}{\log_a b}$.

Hence, $\frac{d^2R_1^{NOMA}}{dp_1^2} < 0$ which implies R_1^{NOMA} is strictly concave.

Similarly, the derivative of (8) with respect to P can be written as,

$$\frac{dR_1^{OFDMA}}{dP} = \frac{\beta}{\log_e 2}\left[\frac{\delta\gamma_{norm}|h_1|^2 2P}{\beta\sigma_1^2 + \delta\gamma_{norm}|Ph_1|^2}\right] \tag{28}$$

The second derivative of R_1^{OFDMA} with respect to P can be written as,

$$\frac{d^2R_1^{OFDMA}}{dP^2} = \frac{\beta}{\log_e 2}\left[\frac{\beta\sigma_1^2 - 2\delta^2\gamma_{norm}^2|P|^2|h_1|^4}{(\beta\sigma_1^2 + \delta\gamma_{norm}|Ph_1|^2)^2}\right], \tag{29}$$

which is always negative if $2\delta^2\gamma_{norm}^2|P|^2|h_1|^4 > \beta\sigma_1^2$. This implies that R_1^{OFDMA} is also strictly concave.

Besides, $\frac{R_1^{NOMA}}{R_1^{OFDMA}}$

$$= \frac{\log_2\left(1 + \frac{|h_{11}|^2p_1}{|h_{12}|^2p_3+\sigma_1^2}\lambda_1\right)}{\beta\log_2\left(1 + \frac{\delta\gamma_{norm}|Ph_1|^2}{\beta\sigma_1^2}\lambda_1\right)} > 1, \quad \text{as } 0 < \beta < 1, \tag{30}$$

and $\frac{\dot{R}_2^{NOMA}}{R_2^{OFDMA}}$

$$= \frac{\log_2\left(1 + \frac{(|h_{21}|^2+|h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3+\sigma_2^2}\lambda_2\right)}{(1-\beta)\log_2\left(1 + \frac{(1-\delta)\gamma_{norm}|Ph_2|^2}{(1-\beta)\sigma_2^2}\lambda_2\right)}$$

$(1-\beta)\dot{R}_2^{NOMA}$

$$= R_2^{OFDMA}\left[\frac{\log_2\left(1 + \frac{(|h_{21}|^2+|h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3+\sigma_2^2}\lambda_2\right)}{\log_2\left(1 + \frac{(1-\delta)\gamma_{norm}|Ph_2|^2}{(1-\beta)\sigma_2^2}\lambda_2\right)}\right]$$

$\dot{R}_2^{NOMA}$

$$= \left[R_2^{OFDMA}\left[\frac{\log_2\left(1 + \frac{(|h_{21}|^2+|h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3+\sigma_2^2}\lambda_2\right)}{\log_2\left(1 + \frac{(1-\delta)\gamma_{norm}|Ph_2|^2}{(1-\beta)\sigma_2^2}\lambda_2\right)}\right] + \beta\dot{R}_2^{NOMA}\right]. \tag{31}$$

Hence, (30) and (31) conclude $R_1^{NOMA} > R_1^{OFDMA}$ $\dot{R}_2^{NOMA} > R_2^{OFDMA}$. ■

V. CAPACITY OPTIMIZATION

The channel capacity of the proposed NOMA system model can be expressed as,

$$C^{NOMA} = \mathbb{E}\left[\log_2\left(1 + \frac{|h_{11}|^2p_1}{\mathbb{E}[|h_{12}|^2]p_3 + \sigma_1^2}\lambda_1\right)\right] + \min\left\{\mathbb{E}\left[\log_2\left(1 + \frac{(|h_{21}|^2 + |h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3 + \sigma_2^2}\lambda_2\right)\right], \mathbb{E}\left[\log_2\left(1 + \frac{(|h_{21}|^2 + |h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{21}|^2]p_1 + \sigma_2^2}\lambda_2\right)\right]\right\}$$

$$+ \mathbb{E}\left[\log_2\left(1 + \frac{|h_{32}|^2p_3}{\mathbb{E}[|h_{31}|^2]p_1 + \sigma_3^2}\lambda_3\right)\right], \tag{32}$$

The channel capacity of cell-1 can be expressed as,

$$\begin{aligned}
C_1^{NOMA} &= \mathbb{E}\left[\log_2\left(1 + \frac{|h_{11}|^2p_1}{\mathbb{E}[|h_{12}|^2]p_3 + \sigma_1^2}\lambda_1\right)\right] \\
&\quad + \mathbb{E}\left[\log_2\left(1 + \frac{(|h_{21}|^2 + |h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3 + \sigma_2^2}\lambda_2\right)\right] \\
&= \mathbb{E}\Bigg[\log_2\left(1 + \frac{|h_{11}|^2p_1}{\mathbb{E}[|h_{12}|^2]p_3 + \sigma_1^2}\lambda_1\right) \\
&\quad + \log_2\left(1 + \frac{(|h_{21}|^2 + |h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3 + \sigma_2^2}\lambda_2\right)\Bigg] \\
&= \mathbb{E}\Bigg[\log_2\Bigg(1 + \frac{|h_{11}|^2p_1}{\mathbb{E}[|h_{12}|^2]p_3 + \sigma_1^2}\lambda_1 + \frac{(|h_{21}|^2+|h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3 + \sigma_2^2}\lambda_2 \\
&\quad + \left(\frac{|h_{11}|^2p_1}{\mathbb{E}[|h_{12}|^2]p_3 + \sigma_1^2}\lambda_1\right)\left(\frac{(|h_{21}|^2 + |h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3 + \sigma_2^2}\lambda_2\right)\Bigg)\Bigg].
\end{aligned} \tag{33}$$

By inequalities $|h_{11}|^2p_1 \gg |h_{12}|^2p_3$ and $|h_{21}|^2 + |h_{22}|^2)\frac{p_2}{2} \gg |h_{22}|^2p_3$, we can confirm $\left(\frac{|h_{11}|^2p_1}{\mathbb{E}[|h_{12}|^2]p_3+\sigma_1^2}\lambda_1\right)\left(\frac{(|h_{21}|^2+|h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3+\sigma_2^2}\lambda_2\right)\right) \geq 1.$, this implies

$$C_1^{NOMA} \geq \mathbb{E}\Bigg[\log_2\Bigg(2 + \frac{|h_{11}|^2p_1}{\mathbb{E}[|h_{12}|^2]p_3 + \sigma_1^2}\lambda_1 + \frac{(|h_{21}|^2 + |h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{22}|^2]p_3 + \sigma_2^2}\lambda_2\Bigg)\Bigg] \tag{34}$$

The channel capacity of cell2 can be expressed as,

$$C_2^{NOMA} = \mathbb{E}\left[\log_2\left(1 + \frac{|h_{32}|^2p_3}{\mathbb{E}[|h_{31}|^2]p_1 + \sigma_3^2}\lambda_3\right)\right] + \mathbb{E}\left[\log_2\left(1 + \frac{(|h_{21}|^2 + |h_{22}|^2)\frac{p_2}{2}}{\mathbb{E}[|h_{21}|^2]p_1 + \sigma_2^2}\lambda_2\right)\right]. \tag{35}$$

which can again be further extended similar to (33).

The channel capacity of the user pair for the OFDMA scheme can be written by

$$\begin{aligned}
C^{OFDMA} &= \beta\log_2\left(1 + \frac{\delta\gamma_{norm}|Ph_1|^2}{\beta\sigma_1^2}\lambda_1\right) \\
&\quad + (1-\beta)\log_2\left(1 + \frac{(1-\delta)\gamma_{norm}|Ph_2|^2}{(1-\beta)\sigma_2^2}\lambda_2\right) \\
&= \log_2\left[\left(1 + \frac{\delta\gamma_{norm}|Ph_1|^2}{\beta\sigma_1^2}\lambda_1\right)^{\beta}\right] \\
&\quad + \log_2\left[\left(1 + \frac{(1-\delta)\gamma_{norm}|Ph_2|^2}{(1-\beta)\sigma_2^2}\lambda_2\right)^{(1-\beta)}\right].
\end{aligned} \tag{36}$$

After the binomial approximation with the truncated series by removing the progressively smaller terms, we can have

$$C^{OFDMA} \approx \log_2\left[\left(1+\frac{\delta\gamma_{norm}|Ph_1|^2}{\sigma_1^2}\lambda_1\right)\right] + \log_2\left[\left(1+\frac{(1-\delta)\gamma_{norm}|Ph_2|^2}{\sigma_2^2}\lambda_2\right)\right]. \quad (37)$$

VI. TRANSMIT POWER OPTIMIZATION

A. POWER ALLOCATION

Even though a lot of work done on power domain NOMA (P-NOMA) in the past, power allocation to the specific UE yet in the place of a challenging issue to be implemented in practice. Here, we have considered the power allocation based on the proposed scheme given in [50]. Two users such as NU and FU with a high SINR difference can jointly work to optimized the NOMA gain, the data symbol associated with NU is assigned a lower transmit power as this UE has the strongest channel gain and vice versa. This implies $p_1 + p_2 = P$, where $p_1 < p_2$.

$$p_1 = \frac{\sqrt{1+\gamma_2}-1}{\gamma_2}, \quad \gamma_2 > 0, \quad (38)$$

where the power allotment factor for NU is defined as a function of the SINR of FU for UE1 transmission.

$$p_2 = P - p_1. \quad (39)$$

B. PROBLEM FORMULATION

To safeguard the QoS of UEs' data transmission, we adopt outage constraint to measure and maintain the high standard network performance. The optimal transmit power design problem with outage probability (OP) constraint under optimal decoding order can be formulated as below,

$$\min_{\Gamma,\mathbf{P},\mathbf{R_c},\mathbf{R_e}} \sum_{i}^{2} P_{\Gamma(i)}, \quad (40)$$

$$\text{subject to: } 0 < P_{\Gamma(i)}, \quad \forall i \in \{1,2\}, \quad (40a)$$

$$R_{e,\Gamma(i)} \le R_{c,\Gamma(i)}, \quad \forall i \in \{1,2\}, \quad (40b)$$

$$R_{c,\Gamma(i)} \le C_{\Gamma(i)}, \quad \forall i \in \{1,2\}, \quad (40c)$$

$$\Theta \le R_{e,\Gamma(i)}, \quad \forall i \in \{1,2\}, \quad (40d)$$

$$\rho_{x_i,\Gamma(i)} \le \zeta, \quad \forall i \in \{1,2\}, \quad (40e)$$

where $\Gamma = [\Gamma(1), \Gamma(2)]$ indicates the decoding vector, $\mathbf{P} = [P_{\Gamma(1)}, P_{\Gamma(2)}]$ indicates power allotment vector, $\mathbf{R_c} = [R_{c,\Gamma(1)}, R_{c,\Gamma(2)}]$ indicates complete data rate vector of the main intelligence with add on redundancy data to safeguard outage constraint, $\mathbf{R_e} = [R_{e,\Gamma(1)}, R_{e,\Gamma(2)}]$ indicates effective data rate vector of the main intelligence, $C_{\Gamma(i)}$ indicates channel capacity of UE $\Gamma(i)$ in order to decode its own intelligence, Θ indicates minimal considerable data rate associated with the useful intelligence, ζ indicates the threshold outage probability of each data symbol x_i and $\rho_{x_i,\Gamma(i)}$ indicates useful outage probability (OP) of data symbol $x_{\Gamma(i)}$ and it can be expressed by

$$\rho_{x_i,\Gamma^*(i)} = \mathbb{P}\{R_{c,\Gamma^*(i)} - R_{e,\Gamma^*(i)} < C_{\Gamma^*(i)}\}, \quad (41)$$

under the optimal condition of (40).

$$R_{c,\Gamma^*(i)} = \begin{cases} \mathbb{E}\left[\log_2\left(1+\frac{\gamma_{w,\Gamma^*(1)}}{\gamma^*_{u,\Gamma(1)}p_3+\sigma_1^2}\lambda_1\right)\right] & \text{if } i=1 \\ \mathbb{E}\left[\log_2\left(1+\frac{\gamma^*_{w,\Gamma(2)}\frac{p_2}{2}}{\gamma^*_{u,\Gamma(2)}p_3+\sigma_2^2}\lambda_2\right)\right] & \text{if } i=2 \end{cases} \quad (42)$$

where $\gamma^*_{w,\Gamma(1)} = \frac{|h^*_{11}|^2}{\sigma_1^2}$, $\gamma^*_{u,\Gamma(1)} = \frac{|h^*_{12}|^2}{\sigma_1^2}$, $\gamma^*_{w,\Gamma(2)} = \frac{(|h^*_{21}|^2+|h^*_{22}|^2)}{\sigma_2^2}$, $\gamma^*_{u,\Gamma(2)} = \frac{|h^*_{22}|^2}{\sigma_2^2}$.

Below the condition provides $R_{e,\Gamma(i)}$ to become optimal,

$$R_{e,\Gamma^*(i)} = \Theta. \quad (43)$$

Thus, (41) can again be re-written as,

$$\rho_{x_i,\Gamma^*(i)} = \mathbb{P}\{R_{c,\Gamma^*(i)} - \Theta < C_{\Gamma^*(i)}\}, \quad (44)$$

Proposition 3: For the optimization problem in (40), the optimal decoding order Γ^* satisfies the following condition,

$$\gamma_{w,\Gamma^*(1)} \ge \gamma_{w,\Gamma^*(2)},$$

where $\Gamma^* = [\Gamma^*(1), \Gamma^*(2)]$ and $\Gamma^*(i) = i, \ \forall i \in \{1,2\}$.

Proof: Proof is delegated in Appendix A. ■

VII. SPECTRAL EFFICIENCY/ENERGY EFFICIENCY OPTIMIZATION

A. POWER CONSUMPTION

Besides, to transmit power, the power is also consumed at the signal processing and the circuit levels, respectively. The total power consumption can be modeled as [51],

$$P_{cons} = \varsigma P_t + P_{bs} + P_{cb}, \quad (45)$$

where the coefficient ς corresponds to the power amplifier, P_t denotes transmit power, the power consume in order to process the signal can be denoted by P_{bs} and P_{cb} represents power consumption due to different circuit blocks.

The power consumption parameter due to baseband signal processing can further be expressed by

$$P_{bs} = W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi], \quad (46)$$

where ϖ is used to include power consumption due to the computation of precoding matrix F, $dim(\mathbf{F})$ represents dimension of F which is used to denote number of RF chains, ε stand for the overhead power consumption at multi-user signal processing level, Ψ indicates power consumption component per unit frequency which does not depend on number of RF chains.

The power consumption at different circuit blocks consists constant components and varying components and can be expressed as follows,

$$P_{cb} = P_{cc} + N_a P_{v1} + R P_{v2} \quad (47)$$

where P_{cc} denotes constant power consumption due to an alternative current to direct current (AC to DC) converter or vice versa at power supply level. A part from varying components, $N_a P_{v1}$ and RP_{v2} are the two different power consumption components and they depend on number of antenna elements and per unit data rate.

Hence again (45) can be expressed by substituting (46) and (47),

$$P_{cons} = \varsigma P_t + W[\varpi [dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cc} + N_a P_{v1} + RP_{v2} \quad (48)$$

B. PROBLEM FORMULATION

The energy efficiency (EE) of any scheme can be done by the ratio of multi-user data rate to total power consumption, which signifies the amount of transferred data bits per unit energy consumed and is measured in [bits/Joule] unit, and as follows,

$$\eta_{EE} = \begin{cases} \dfrac{R^{NOMA}}{P_{cons}} & \text{for NOMA scheme} \\ \dfrac{R^{OFDMA}}{P_{cons}} & \text{for OFDMA scheme} \end{cases} \quad (49)$$

The spectral efficiency (SE) signifies the amount of transferred data bits per unit bandwidth and is measured in [bits/s/Hz] unit, and as follows,

$$\eta_{SE} = \begin{cases} \dfrac{R^{NOMA}}{W} & \text{for NOMA scheme} \\ \dfrac{R^{OFDMA}}{W} & \text{for OFDMA scheme} \end{cases} \quad (50)$$

The problem formulation is performed in order to achieve optimum EE be subjected to a necessary minimum SE, denoted by $\check{\eta}_{SE}$, needs. As $\frac{d\eta_{SE}}{dP} > 0$, this implies that η_{SE} increases with P. Therefore, lower boundary of SE can be obtained from with the minimum transmit power $\check{P}$. Hence, $\eta_{SE}\Big|_{P=\check{P}} = \check{\eta}_{SE}$, which implies $\check{R} = \check{\eta}_{SE} W$. In general, considering all the constraints, we formulate the EE optimization problem for NOMA scheme as,

$$\begin{cases} \max_{\pi,\mathbf{P}} \eta_{EE} = \max_{\pi,\mathbf{P}} \dfrac{R^{NOMA}(p_1, p_2)}{P_{cons}} \\ \max_{\pi,\mathbf{P}} \eta_{SE} = \max_{\pi,\mathbf{P}} \dfrac{R^{NOMA}(p_1, p_2)}{W} \end{cases} \quad (51)$$

$$\text{Subject to: } 0 < P \le P_{opt} \quad (51a)$$
$$R^{NOMA}(p_1, p_2) \ge \check{R}^{NOMA}(p_1, p_2) \quad (51b)$$
$$R_i \ge R_i^{min}, \quad \forall i = \{1, 2\} \quad (51c)$$
$$\check{R}^{NOMA}(p_1, p_2) \ge R_1^{min} + R_2^{min} \quad (51d)$$
$$P_{opt} \ge \check{P} \quad (51e)$$
$$0 < \pi < 0.5 \quad (51f)$$

where $\pi = \frac{p_1}{P}$ denotes power splitting factor.

Proposition 4: The objective function given in (49) as η_{EE} is a strictly quasi-concave concerning P irrespective of NOMA scheme or OFDMA scheme.

Proof: Proof is delegated in Appendix B. ■

Proposition 5: Smaller amount of circuit power consumption can provide better optimal η_{EE} at smaller transmit power irrespective of $\check{P}$ and P_{max}.

Proof: Proof is delegated in Appendix C. ■

C. SE-EE TRADE-OFF

In this subsection, we establish a SE-EE trade-off framework by applying the weighted sum method to deal with multi-objective function, $\eta_{MOF}(\pi, P) = [\eta_{SE}(\pi, P), \eta_{EE}(\pi, P)]^T$. The SE-EE trade-off is generally consisting of all accomplishable (η_{SE}, η_{EE}) pairs. The multi-objective optimization problem in (51) can be transformed into the following equivalent expression employing the weighted sum method,

$$\max_{\pi,\mathbf{P}} \Phi \Upsilon_{EE} \eta_{EE} + (1 - \Phi) \Upsilon_{SE} \eta_{SE}, \quad (52)$$

$$\text{Subject to: } \sum_{i=1}^{2} R_i \ge R_i^{min}, \quad \forall i = \{1, 2\} \quad (52a)$$
$$\sum_{i=1}^{2} p_i \le P_{max}, \quad \forall i = \{1, 2\} \quad (52b)$$
$$\mathbb{E}[|h_{12}|^2] p_3 \le I_i^{th}, \quad i = 1 \quad (52c)$$
$$\mathbb{E}[|h_{22}|^2] p_3 \le I_i^{th}, \quad i = 1 \text{ at UE1} \quad (52d)$$
$$p_i \ge 0, \quad \forall i = \{1, 2\} \quad (52e)$$
$$\pi \le 1 \quad (52f)$$

where Υ_{EE} and Υ_{SE} are normalization factors[2] considered to bring into the same range of the objective functions due to unequal amount of bandwidth and transmit power allocation to a higher degree, Φ stand for trade-off parameter that signifies the objective function's priority and its range lies as follows $0 \le \Phi \le 1$.

Thus, (52) can be re-expressed to extend its analysis for the rest of this sub-section as below,

$$\max_{\pi,\mathbf{P}} \Upsilon_{EE} \eta_{EE} + \left(\frac{1 - \Phi}{\Phi}\right) \Upsilon_{SE} \eta_{SE}, \quad (53)$$

$$\text{Further, } \Lambda = \max_{\pi,\mathbf{P}} \eta_{EE} + \vartheta \left(\frac{\Upsilon_{SE} \eta_{SE}}{\Upsilon_{EE}}\right), \quad \forall \vartheta = \left(\frac{1 - \Phi}{\Phi}\right)$$
$$= \max_{\pi,\mathbf{P}} \eta_{EE} \left[1 + \vartheta \left(\frac{\Upsilon_{SE} \eta_{SE}}{\Upsilon_{EE} \eta_{EE}}\right)\right]$$
$$\text{Subject to: } (52a) - (52f), (54a);$$
$$\vartheta \ge 0, (54b) \quad (54)$$

Proposition 6: Λ in (54) as is a quasi-concave with respect to P irrespective of NOMA scheme or OFDMA scheme.

Proof: Proof is delegated in Appendix D. ■

VIII. PERFORMANCE COMPARISON

This section gives a comprehensive comparison of OMA/NOMA 5G networks from the perspective of user data rates, capacity, SE, EE, and so on. Here, note that proposed NOMA scheme is assisted by ICC whereas CCNOMA scheme is assisted by SIC.

[2] Υ_{EE} and Υ_{SE} depend on the network parameters, for example total power budget.

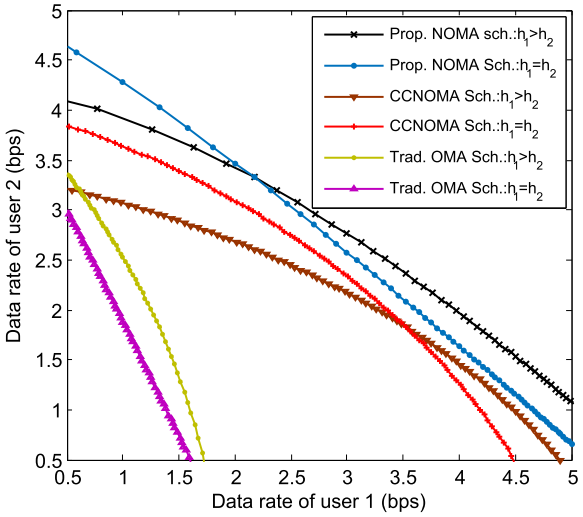


FIGURE 4. **Performance comparisons of three schemes at different channel conditions for user 2's data rate and user 1's data rate trade-off.**

A. IMPROVED USER DATA RATES AND CAPACITY

In 4G cellular networks, OMA has been extensively considered to reach a higher data rate. Mobile data traffic is expected to grow more than 1000 times in 2030 than that in 2020. Therefore, network capacity needs to be expanded to satisfy the massive demand for data rates. The capacity scaling for 5G is dealt with by advanced multi-user techniques such as NOMA as a probable alternative to OMA. In this regard, the role of different technology enablers, which include MIMO, mmWave communication, usage of unlicensed spectrum, network densification by small cell deployment cannot be ignored.

Figure 4 shows a trade-off between the data rate of user 1 and the data rate of user 2, where the NOMA scheme offers the data rate gain response similar to the OMA scheme to a great extent, with an average gain of 28.12% for $h_1 > h_2$ and 21.1% for $h_1 = h_2$ at optimum point, respectively. The variation from maxima to minima for both schemes is approximately 28.12-50% for $h_1 > h_2$ and 21.1-21.4% for $h_1 = h_2$, respectively. The results of figures 5(a) and 5(b) for all the schemes are showing an indication that the NOMA scheme is much suitable for the channel condition $h_1 > h_2$, whereas OMA is achieving better rate gain for $h_1 = h_2$ compared to $h_1 > h_2$. For OFDMA scheme, the data rate is penalyzed to a notably large extent due to the operation of orthogonal resource allocation, the alignment of RBs to the symbol and the cyclic prefix (CP) length to each of the users, respectively. In comparison to OFDMA, user 2 can secure a higher data rate without lowering the user 1's data rate for the NOMA schemes, although proposed NOMA scheme approaches the optimal response. The NOMA schemes are providing perfect balance on the data rate optimization of two users, whereas OMA scheme can be compromising with the lower data rate of user 1 to optimize user 2's data rate.

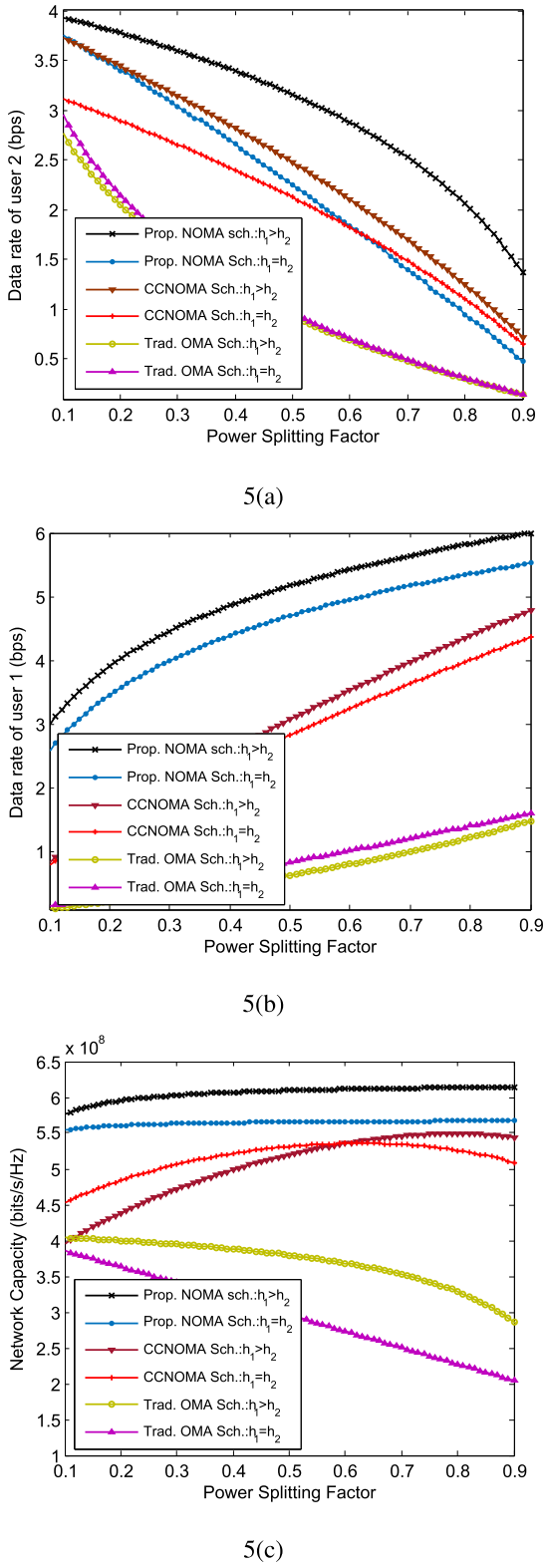


FIGURE 5. **Performance comparisons of three schemes at different channel conditions, where (a) demonstrates user 2's data rate as a function power splitting factor (b) demonstrates user 1's data rate as a function of power spilitting factor and (c) demonstrates network capacity as a function of power spilitting factor.**

Furthermore, this statement validates the Proposition 2 and the derived closedform expressions therein.

Figures 5(a) and 5(b) compare two users' optimum data rates obtained by a NOMA scheme to that obtained by an OMA scheme for varying power splitting factors. User 1 who has a strong channel condition compared to other users is capable of exerting full liberty in a NOMA transmission principle whereas the same user is capable of exerting partial liberty in an OMA transmission principle, hence the data rate of user 1 on every occasion is higher for NOMA than OMA if varying power splitting factor follows less than 1 condition. From figure 5(a) we can also observe that user 2 in OMA performs better than NOMA in terms of data rate irrespective of the difference of the channel conditions for two users are significant or not. This is due to the reason that the strong user signal is becoming interfere in decoding the weak user signal according to the NOMA principle, which in turn limits the user 2's data rate. Fairness index ranges between 0 and 1. The data rate allocation to the user is ideally fair if $J = 1$. Generally, J reduces with the increase density of the BS deployment for a given number of RBs. This is due to the overlapping of cells' network coverage which causes the aggregated interference to the users. However, the users with worst channel condition may not be evaluated by network because of resource scarcity. (19) tells us that J increases with the increase of data rate. NOMA's potential to multiplexing large BSs on each RB produces higher amount of multi-user diversity gain, which is the reason of achieving improved fairness level in the NOMA schemes, although proposed NOMA scheme outperforms CCNOMA scheme.

Based on analytical work of the multi-user capacity, figure 5(c) illustrates the comparison of OMA and NOMA for network capacity, where two users in the additive white Gaussian noise (AWGN) channel are assumed with Monte-Carlo simulations. In the AWGN channel with intersymbol interference, even though OMA could accomplish the network capacity in the downlink but NOMA is optimum whilst OMA is strictly sub-optimum if channel state information (CSI) is available to the receiver only. Hence, the saturation level of the NOMA network capacity for power splitting factor more than 0.3 is constant but considerably higher for the NOMA scheme compared to the OMA scheme. In comparison with traditional OMA scheme, CCNOMA and proposed NOMA scheme provide a worthy of attention network capacity gain, as NOMA permits the users to share the same spectrum and spatial resource blocks, and thus surpasses the performance. In addition, we can observe that the performance obtained by the proposed NOMA scheme is superior to other two schemes, which is due to the application of ICC; while OFDMA becomes inferior to MRC detection in the higher splitting factor regime.

B. SE AND EE TRADE-OFF

As illustrated in figure 6, six plots are simulated with the x-axis being the SE and the y-axis being the EE, respectively, of the networks for two users' scenario. The disparity of

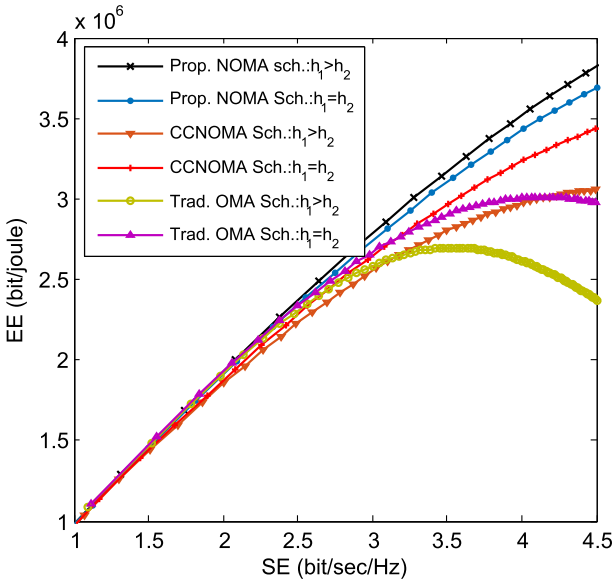
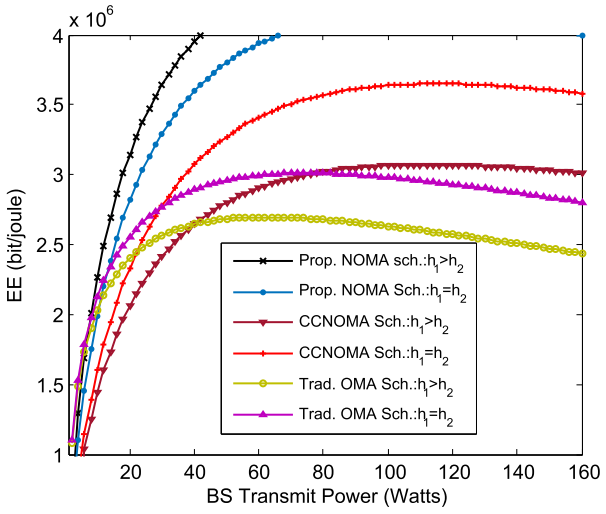


FIGURE 6. Performance comparisons of three schemes at different channel conditions for SE-EE trade-off.

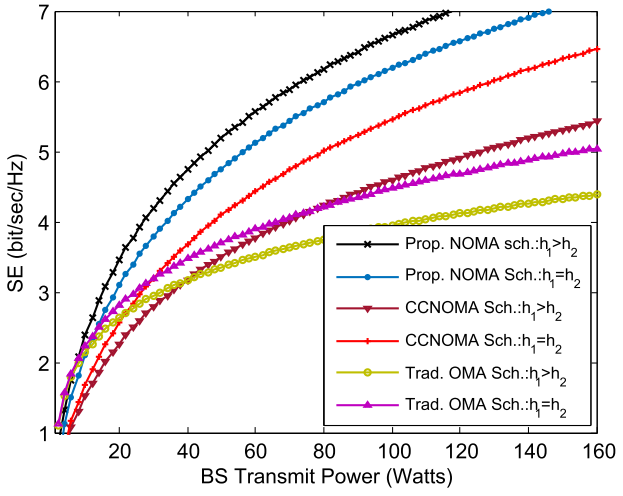
channel fading between the two users has a strong effect in OMA compared to NOMA, this implies much better EE-SE performance of NOMA than OMA in the scenarios which are more realistic for practical. However, the NOMA scheme outperforms the OMA scheme in the upper SE regime and there is not much performance difference that can be noticed at low to medium SE regimes. Besides, a large gap occurs between the two assumed channel conditions in each scheme in the upper SE regime, e.g. 3.4×10^6 bit/Joule at 4.5 bps/Hz for $h_1 > h_2$ vs. 3.1×10^6 bit/Joule at 4.5 bps/Hz for $h_1 = h_2$ in NOMA scheme respectively and 2.4×10^6 bit/Joule at 4.5 bps/Hz for $h_1 > h_2$ vs. 3×10^6 bit/Joule at 4.5 bps/Hz for $h_1 = h_2$ in OMA scheme respectively. It is further shown by the trade-off curves of EE and SE that with the increase of SE from 1 bps/Hz to 3.6 bps/Hz, the EE also increases and reaches its optimum value in the OMA scheme for $h_1 > h_2$. And if SE continues to increase, the EE will deteriorate, particularly in the OMA scheme irrespective of channel disparity. This can be seen from the figure that when $\eta_{SE} \leq 2.75\, bit/sec/Hz$, three schemes achieve almost the same EE; however the proposed scheme continues to improve the EE with further increase of η_{SE}, and outperforms CCNOMA and traditional OMA schemes respectively. This confirms proposition 6 as well.

C. ENERGY EFFICIENCY (EE)

On top of everything else EE has got importance as the information and communication technology (ICT) consumes about 15% of the total world energy dissipation and this is becoming a prime issue worldwide to society and humanity at large. The resource allocation in DL-NOMA is performed by articulating on EE and assume bits/Joule to evaluate EE accomplishment. The key problems in subchannel allotment



7(a)



7(b)

FIGURE 7. Performance comparisons of three schemes at different channel conditions, where (a) demonstrates EE as a function of BS transmit power and (b) demonstrates SE as a function of BS transmit power.

and power assignment have been formulated by decoupling them subject to an EE optimization constraint. Here, one subchannel can multiplex up to two users which results in ease of SIC computations. However, the EE metric is stated as the ratio of the users' sum data rate to the total power consumption. Figure 7(a) illustrates the EE of the networks as a function of the BS transmit power. It can be noticed that the energy consumption reduces as the transmit power rises and gradually converges to a stable value, which validates proposition 4. More precisely, the proposed scheme has a much better EE performance compared to the other two schemes due to minimum circuit power consumption at par proposition 5, when the transmit power is more than 20 Watts. It can also be observed that the performance gap of the EE among three schemes is relatively large when transmit power lies in the middle range and the gap does not increase, infact decreases as the transmit power further increases, which again validates proposition 5.

D. SPECTRAL EFFICIENCY (SE)

NOMA is an optimum scheme of utilizing spectrum in both UL and DL transmissions, respectively. This is due to the reason that each NOMA user is exploiting the complete bandwidth, whereas OMA users are restricted to a partial part of the bandwidth and it decreases at the same rate that the number of users increases. Besides, NOMA is well capable to be an integral part of the systems with other technology enablers, such as MIMO, beamformer, cluster, and mmWave, over a greater expanse of space or time to achieve even more throughput. Due to the given advantages, NOMA has brought special attention from the fields of academia, research and innovation, and industry. As shown in figure 7(b), when transmit power is larger than 60 watts, the SE of the proposed NOMA scheme improves significantly and becomes superior to both CCNOMA and OMA schemes.

IX. CONCLUSION

In this paper, we discussed the performance comparison between NOMA and OMA schemes respectively, and focused on the key problem formulation for the optimization of user data rates, capacity, transmit power, SE and EE. The direction toward the possible new analytical framework for NOMA from OMA undergoes continual evolution in a paradigm shift of 4G to 5GB cellular technologies. In particular, we have proposed a novel NOMA scheme with its special transceiver design and have shown that it works very well for the promising multi-objective optimization framework. The simulation results have also confirmed a standard of excellence for the proposed scheme over CCNOMA and OFDMA while sustaining a reasonable level of QoS.

APPENDIX A
PROOF OF PROPOSITION 3

For an unknown decoding order, let us we consider $\Gamma(l) = a$ and $\Gamma(l+1) = b$ with $\gamma_a > \gamma_b$ and power allotment vector $\mathbf{P} = [P_{\Gamma(1)}, P_{\Gamma(2)}]$. If we exchange the decoding order of UEs with $\Gamma(l) = b$ and $\Gamma(l+1) = a$, then the following metric $\gamma_{w,\Gamma(i)} = \min_{i\in\{1,2\}} \gamma_{\Gamma(i)}$, $i \neq (l+1)$ is remained unalter, whereas $\gamma_{w,\Gamma(l+1)}$ is either becoming greater or remain unalter. Hence by following (44), we can confirm that $\rho_{x_i,\Gamma^*(i)}$ is becoming smaller with the increase of $\gamma_{w,\Gamma(i)}$. Therefore, exchanging them does not effect the useful OP of any data symbol $x_{\Gamma(i)}$, $i \neq (b+1)$, whereas the OP of the data symbol $x_{\Gamma(b+1)}$ may become smaller. This also implies that the decreased or unaltered transmission power given in (40) does not put impact on outage constraint. Hence, exchanging the decoding order of the given user pair has no effect on the performance of the proposed scheme. Hence, the computation by iteratively exchanging Γ of the given user

pair for optimal decoding order is to be performed till we can be achieved with $\gamma_{w,\Gamma(i)}$ in descending order.

APPENDIX B
PROOF OF PROPOSITION 4

We re-express the objective function given in (49) as,

$$\eta_{EE} = \frac{R^{NOMA}}{P_{Cons}} = \frac{R_1^{NOMA}(p_1) + R_2^{NOMA}(p_2)}{(\varsigma P_t + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cc} + N_a P_{v1} + RP_{v2})} = \frac{R^{OFDMA}}{P_{Cons}} = \frac{R_1^{OFDMA}(P) + R_2^{OFDMA}(P)}{(\varsigma P_t + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cc} + N_a P_{v1} + RP_{v2})} \quad \text{(B.1)}$$

Therefore, the first derivative of (B.1) with respect to P can be written as in (B.2), as shown at the bottom of the page.

As $\frac{dR_1^{NOMA}(P)}{dP} > 0$ and $\frac{d^2R_1^{NOMA}(P)}{dP^2} < 0$, we will have $\frac{dR^{NOMA}(P,0)}{dP} > 0$ and $\frac{d^2R^{NOMA}(P,0)}{dP^2} < 0$. Similarly, as $\frac{dR_1^{OFDMA}(P)}{dP} > 0$, $\frac{dR_2^{OFDMA}(P)}{dP} > 0$, $\frac{d^2R_1^{OFDMA}(P)}{dP^2} < 0$ and $\frac{d^2R_2^{OFDMA}(P)}{dP^2} < 0$, we will have $\frac{dR^{OFDMA}(P)}{dP} > 0$ and $\frac{d^2R^{OFDMA}(P)}{dP^2} < 0$. This implies that both $R^{NOMA}(P,0)$ and $R^{OFDMA}(P)$ are strictly concave function of P. This confirms that η_{EE} is a strictly concave function of P with its affine denominator and this proves η_{EE} is a strictly quasi-concave with respect to P irrespective of NOMA scheme or OFDMA scheme.

APPENDIX C
PROOF OF PROPOSITION 5

First of all, this is evident from (49) that η_{EE} reduces with the increase of P_{cb}. Let us consider two different circuit power consumption level of P_{cb} with the assumption $P_{cb1} < P_{cb2}$. In one condition it consists static component only with the notation P_{cb1}, whereas in another condition it consists both static and dynamic components with the notation P_{cb2}. We also consider the notations $\eta_{EE(cb1)}$ and $\eta_{EE(cb2)}$ against P_{cb1} and P_{cb2}, respectively. Besides, we assume two points P_{opt1} and P_{opt2} in $x-$axis at which EE reaches to its maximum value for the respective case studies. As $P_{cb1} < P_{cb2}$, hence we achieve $\eta_{EE(cb1)} > \eta_{EE(cb2)}$.

Based on the maxima condition, we can write the following as

$$\left.\frac{d\eta_{EE(cb1)}}{dP}\right|_{P_t=P, P=P_{opt1}} = 0. \quad \text{(C.1)}$$

By following (B.2), (C.1) can further extend into (C.2), as shown at the bottom of the page.

Therefore, (C.3) and (C.4), as shown at the bottom of the page, signifies that the function $\eta_{EE(cb1)}$ is increasing concavely from $\check{P}$ to P_{opt1}.

$$\frac{d\eta_{EE}}{dP} \approx \left[\frac{\left(\frac{dR_1^{NOMA}(dP)}{dP} + \frac{dR_2^{NOMA}(0)}{dP}\right)(\varsigma P + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cc} + N_a P_{v1} + RP_{v2}) - \left(\varsigma + P_{v1}\left(\frac{dN_a}{dP}\right) + P_{v2}\left(\frac{dR_1^{NOMA}(P)}{dP} + \frac{dR_2^{NOMA}(0)}{dP}\right)\right)R^{NOMA}(P,0)}{(\varsigma P + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cc} + N_a P_{v1} + RP_{v2})^2}\right]\Bigg|_{P_t=P,\, p_1=P}$$
$$\approx \left[\frac{\left(\frac{dR_1^{OFDMA}(P)}{dP} + \frac{dR_2^{OFDMA}(P)}{dP}\right)(\varsigma P + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cc} + N_a P_{v1} + RP_{v2}) - \left(\varsigma + P_{v1}\left(\frac{dN_a}{dP}\right) + P_{v2}\left(\frac{dR_1^{OFDMA}(P)}{dP} + \frac{dR_2^{OFDMA}(P)}{dP}\right)\right)R^{OFDMA}(P)}{(\varsigma P + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cc} + N_a P_{v1} + RP_{v2})^2}\right]\Bigg|_{P_t=P} \quad \text{(B.2)}$$

$$\left[\frac{\left.\frac{dR^{NOMA}(P,0)}{dP}\right|_{P=P_{opt1}} (\varsigma P_{opt1} + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cb1}) - \left(\varsigma + P_{v1}\left(\frac{dN_a}{dP}\right)\right)R^{NOMA}(P,0)}{(\varsigma P_{opt1} + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cb1})^2}\right]\Bigg|_{P_{cb1}=P_{cc}} = 0. \quad \text{(C.2)}$$

$$\text{Now, } \left.\frac{d\eta_{EE(cb2)}}{dP}\right|_{P_t=P, P=P_{opt1}} > 0, \quad \forall P \in [\check{P}, P_{opt1}] \quad \text{(C.3)}$$

$$\text{Since, } \left[\frac{\left.\frac{dR^{NOMA}(P,0)}{dP}\right|_{P=P_{opt1}} (\varsigma P_{opt2} + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cb2}) - \left(\varsigma + P_{v1}\left(\frac{dN_a}{dP}\right) + P_{v2}\left(\left.\frac{dR^{NOMA}(P,0)}{dP}\right|_{P=P_{opt1}}\right)\right)R^{NOMA}(P,0)}{(\varsigma P_{opt2} + W[\varpi[dim(\mathbf{F})]^{\varepsilon+1} + \Psi] + P_{cb2})^2}\right]\Bigg|_{P_{cb2}=P_{cc}+N_aP_{v1}+RP_{v2}} > 0 \quad \text{(C.4)}$$

However, we find the maxima of $\eta_{EE(cb2)}$ at

$$\left.\frac{d\eta_{EE(cb2)}}{dP}\right|_{P_t=P,P=P_{opt2}} = 0. \tag{C.5}$$

This proofs $P_{opt2} > P_{opt1}$. Thus it is very crucial to maintain minimum power consumption at different circuit blocks to design an energy-efficient system model.

APPENDIX D
PROOF OF PROPOSITION 6

(54) can further extend to the following expression to prove that Λ is a quasi-concave with respect to P,

$$\begin{aligned}\Lambda &= \max_{\pi,\mathbf{P}} \frac{R^{NOMA}}{P}\left[1+\vartheta\left(\frac{\Upsilon_{SE}P}{\Upsilon_{EE}W}\right)\right] \\ &= \max_{\pi,\mathbf{P}} \frac{R^{NOMA}}{P}\left(1+\Delta P\right) \\ &= \max_{\pi,\mathbf{P}} \left\{\frac{R^{NOMA}}{P}+\Delta R^{NOMA}\right\} \\ &= \max_{\pi,\mathbf{P}} \left\{\eta_{EE}+\Delta R^{NOMA}\right\}, \end{aligned} \tag{D.1}$$

where $\Delta = \vartheta\left(\frac{\Upsilon_{SE}}{\Upsilon_{EE}W}\right)$.

In Appendix B, it is conclusive that both $R^{NOMA}(P,0)$ and $R^{OFDMA}(P)$ are a strictly concave function of P. From proposition 4, we also observe that η_{EE} is a strictly quasi-concave for P regardless of the NOMA scheme or OFDMA scheme. As Λ is a linear combination or summation of two segments such as η_{EE} and ΔR^{NOMA}, thus it implies that Λ is becoming a strictly quasi-concave concerning P.

REFERENCES

[1] D. Yuan, J. Joung, C. K. Ho, and S. Sun, "On tractability aspects of optimal resource allocation in OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 863–873, Feb. 2013.

[2] L. Lei, D. Yuan, C. K. Ho, and S. Sun, "A unified graph labeling algorithm for consecutive-block channel allocation in SC-FDMA," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5767–5779, Nov. 2013.

[3] J. Li, X. Lei, P. D. Diamantoulakis, F. Zhou, P. Sarigiannidis, and G. K. Karagiannidis, "Resource allocation in buffer-aided cooperative non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7429–7445, Dec. 2020.

[4] Z. Shi, W. Gao, S. Zhang, J. Liu, and N. Kato, "Machine learning-enabled cooperative spectrum sensing for non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5692–5702, Sep. 2020.

[5] J. Wang, X. Kang, S. Sun, and Y.-C. Liang, "Throughput maximization for peer-assisted wireless powered IoT NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5278–5291, Aug. 2020.

[6] I. Baig, "A precoding-based multicarrier non-orthogonal multiple access scheme for 5G cellular networks," *IEEE Access*, vol. 5, pp. 19233–19238, 2017.

[7] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[8] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "The impact of power allocation on cooperative non-orthogonal multiple access networks with SWIPT," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4332–4343, Jul. 2017.

[9] J. Choi, "Random access with layered preambles based on NOMA for two different types of devices in MTC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 871–881, Feb. 2021.

[10] X. Yue, Y. Liu, S. Kang, A. Nallanathan, and Z. Ding, "Exploiting full/half-duplex user relaying in NOMA systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 560–575, Feb. 2017.

[11] J. Men, J. Ge, and C. Zhang, "Performance analysis of nonorthogonal multiple access for relaying networks over Nakagami-*m* fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1200–1208, Feb. 2017.

[12] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654–667, Feb. 2016.

[13] Z. Yang, Y. Liu, Y. Chen, and N. Al-Dhahir, "Cache-aided NOMA mobile edge computing: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6899–6915, Oct. 2020.

[14] Y. Liu, H. Yu, S. Xie, and Y. Zhang, "Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11158–11168, Nov. 2019.

[15] J. Chen, S. Chen, Q. Wang, B. Cao, G. Feng, and J. Hu, "iRAF: A deep reinforcement learning approach for collaborative mobile edge computing IoT networks," *IEEE Int. Things J.*, vol. 6, no. 4, pp. 7011–7024, Aug. 2019.

[16] Z. Ding, J. Xu, O. A. Dobre, and V. Poor, "Joint power and time allocation for NOMA-MEC offloading," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6207–6211, Mar. 2019.

[17] W. Wu, F. Zhou, R. Q. Hu, and B. Wang, "Energy-efficient resource allocation for secure NOMA-enabled mobile edge computing networks," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 493–505, Jan. 2020.

[18] Z. Zhang, Z. Ma, Y. Xiao, M. Xiao, G. K. Karagiannidis, and P. Fan, "Non-orthogonal multiple access for cooperative multicast millimeter wave wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1794–1808, Aug. 2017.

[19] Z. Zhang, H. Sun, R. Q. Hu, and Y. Qian, "Stochastic geometry based performance study on 5G non-orthogonal multiple access scheme," in *Proc. IEEE GLOBECOM*, Dec. 2016, pp. 1–6.

[20] Z. Xiang, W. Yang, G. Pan, Y. Cai, and Y. Song, "Physical layer security in cognitive radio inspired NOMA network," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 700–714, Jun. 2019.

[21] P. Swami, M. K. Mishra, V. Bhatia, and T. Ratnarajah, "Performance analysis of NOMA enabled hybrid network with limited feedback," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4516–4521, Apr. 2020.

[22] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[23] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. 77th IEEE Veh. Technol. Conf.*, Dresden, Germany, Jun. 2013, pp. 1–5.

[24] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[25] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "On the sum rate of MIMO-NOMA and MIMO-OMA systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 534–537, Aug. 2017.

[26] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "Capacity comparison between MIMO-NOMA and MIMO-OMA with multiple users in a cluster," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2413–2424, Oct. 2017.

[27] M. Zeng, W. Hao, O. A. Dobre, and H. V. Poor, "Energy-efficient power allocation in uplink mmWave massive MIMO with NOMA," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3000–3004, Mar. 2019.

[28] M. Hedayati and I.-M. Kim, "On the performance of NOMA in the two-user SWIPT system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11258–11263, Nov. 2018.

[29] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, Sep. 2016.

[30] Y. Zhang, H.-M. Wang, T.-X. Zheng, and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2852–2857, Mar. 2017.

[31] M. Morales-Cespedes, O. A. Dobre, and A. Garcia-Armada, "Semi-blind interference aligned NOMA for downlink MU-MISO systems," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1852–1865, Mar. 2020.

[32] P. Kheirkhah Sangdeh, H. Pirayesh, Q. Yan, K. Zeng, W. Lou, and H. Zeng, "A practical downlink NOMA scheme for wireless LANs," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2236–2250, Apr. 2020.

[33] S. Silva, G. A. A. Baduge, M. Ardakani, and C. Tellambura, "NOMA-aided multi-way massive MIMO relaying," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4050–4062, Jul. 2020.
[34] X. Ge, H. Cheng, M. Guizani, and T. Han, "5G wireless backhaul networks: Challenges and research advances," *IEEE Netw.*, vol. 28, no. 6, pp. 6–11, Nov./Dec. 2014.
[35] J. Zhao, T. Q. S. Quek, and Z. Lei, "Heterogeneous cellular networks using wireless backhaul: Fast admission control and large system analysis," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2128–2143, Oct. 2015.
[36] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless backhauling of 5G small cells: Challenges and solution approaches," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 22–31, Oct. 2015.
[37] X. Zou, B. He, and H. Jafarkhani, "On uplink asynchronous non-orthogonal multiple access systems with timing error," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.
[38] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6884–6898, Oct. 2020.
[39] J. Zhu, Y. Huang, J. Wang, K. Navaie, and Z. Ding, "Power efficient IRS-assisted NOMA," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 900–913, Feb. 2021.
[40] F. Fang, Y. Xu, Q. Pham, and Z. Ding, "Energy-efficient design of IRS-NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 14088–14092, Nov. 2020.
[41] J. Xu, J. Li, S. Gong, K. Zhu, and D. Niyato, "Passive relaying game for wireless powered Internet of Things in backscatter-aided hybrid radio networks," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8933–8944, Oct. 2019.
[42] W. Chen, H. Ding, S. Wang, D. B. da Costa, F. Gong, and P. H. J. Nardelli, "Backscatter cooperation in NOMA communications systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3458–3474, Jun. 2021, doi: 10.1109/TWC.2021.3050600.
[43] M. Bennis, M. Debbah, and H. V. Poor, "Ultra-reliable and low-latency wireless communication: Tail, risk and scale," *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.
[44] A. Anand and G. de Veciana, "Resource allocation and HARQ optimization for URLLC traffic in 5G wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2411–2421, Nov. 2018.
[45] H. Zhang, Y. Qiu, K. Long, G. K. Karagiannidis, X. Wang, and A. Nallanathan, "Resource allocation in NOMA-based fog radio access networks," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 110–115, Jun. 2018.
[46] X. Wen, H. Zhang, H. Zhang, and F. Fang, "Interference pricing resource allocation and user-subchannel matching for NOMA hierarchy fog networks," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 467–479, Jun. 2019.
[47] I. Randrianantenaina, M. Kaneko, H. Dahrouj, H. ElSawy, and M.-S. Alouini, "Interference management in NOMA-based fog-radio access networks via scheduling and power allocation," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5056–5071, Aug. 2020.
[48] L. Wang, E. Sasoglu, B. Bandemer, and Y.-H. Kim, "A comparison of superposition coding schemes," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2970–2974.
[49] R. K. Jain, D. M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Eastern Res. Lab., Digit. Equip. Corp., Maynard, MA, USA, Tech. Rep. DEC-TR-301, 1984.
[50] Y. Zhu, H. J. E. Kwon, H. Jung, U. Kumar, and J.-K. J. K. Fwu, "Nonorthogonal multiple access (NOMA) wireless systems and methods," U.S. Patent EP3 138 227 A1, Mar. 8, 2017.
[51] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 954–965, Oct. 2014.
[52] C. Eklund, R. B. Marks, K. L. Stanwood, and S. Wang, "IEEE standard 802.16: A technical overview of the WirelessMAN air interface for broadband wireless access," *IEEE Commun. Mag.*, vol. 40, no. 6, pp. 98–107, Jun. 2002.
[53] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.