IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 31 October 2023, accepted 20 November 2023, date of publication 28 November 2023,
date of current version 6 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3337536

RESEARCH ARTICLE

A Novel Transceiver and an Asynchronous Mode for the Hybrid Multiple-Access HetNet Architecture

JOYDEV GHOSH[1], CESAR VARGAS-ROSALES[2], (Senior Member, IEEE), LUCIANO LEONEL MENDES[3], (Member, IEEE), IN-HO RA[4], (Member, IEEE), VAN NHAN VO[5,6], PHET AIMTONGKHAM[1], AND CHAKCHAI SO-IN[1], (Senior Member, IEEE)

[1]Applied Network Technology, Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand
[2]School of Engineering and Science, Tecnológico de Monterrey, Monterrey 64849, Mexico
[3]National Telecommunications Institute—Inatel, Santa Rita do Sapucai 37540-000, Brazil
[4]School of Computer, Information and Communication Engineering, Kunsan National University, Gunsan 54150, South Korea
[5]Faculty of Information Technology, Duy Tan University, Da Nang 550000, Vietnam
[6]Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

Corresponding author: Chakchai So-In (chakso@kku.ac.th)

This work was supported in part by the College of Computing, Khon Kaen University; in part by Khon Kaen University; and in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government under Grant 2021R1A2C2014333.

ABSTRACT Multiuser gigantic-multiple-input multiple-output (MU-gMIMO) and nonorthogonal multiple access (NOMA) are jointly seen as important enabling technologies for sixth generation (6G) networks. They have many benefits, such as spatial multiplexing, spatial diversity, massive connectivity, and spectral efficiency (SE). However, gMIMO-NOMA suffers from many inherent challenges. In this paper, we propose an MU-gMIMO-hybrid multiple-access (HMA) heterogeneous network architecture to address the 'nearly same channel gain' issue. Then, an iterative minimal mean squared error (IMMSE) scheme is applied along with quadrature amplitude modulation (QAM) for maximal ratio transmission in the proposed transceiver design to address the 'residual error' caused by imperfect successive interference cancellation (ISIC). Finally, to assess the performance of the proposed architecture for the 'time offsets' issue, we investigate the asynchronous mode with an MMSE detector matrix following imperfect channel state information (ICSI) to provide a new analysis for HMA transmission and formulate an optimization problem for energy efficiency (EE).

INDEX TERMS Gigantic-multiple-input multiple-output (gMIMO), hybrid multiple access (HMA), asynchronous transmission, quadrature amplitude modulation (QAM), iterative interference cancellation (IIC), spectral efficiency (SE), energy efficiency (EE), sixth-generation (6G).

I. INTRODUCTION

Different orthogonal multiple access (OMA) techniques, such as frequency-division multiple access (FDMA), time-division multiple access (TDMA), code-division multiple access (CDMA), and orthogonal frequency-division multiple access (OFDMA), have given each generation of wireless networks—from the first generation (1G) to the fourth generation (4G)—distinctive features. In the early stages of beyond 4G research, non-orthogonal multiple access (NOMA) is acknowledged as having more potential than OMA. But since 5G also uses OFDMA, NOMA techniques haven't completely taken over from the fourth to the fifth generation. The sixth generation (6G) of mobile networks will require a cutting-edge technology that enables the IMT-2030 vision to be achieved [1]. To create new research opportunities over the following few decades, it is now especially important to identify the operational features of 5G networks, including customer demands, key drive options, open issues, and future challenges.

Users are permitted to use resources in an orthogonal manner under classic orthogonal multiple access (OMA),

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

which causes severe congestion when numerous users attempt to access a resource at once. In this context, nonorthogonal multiple access (NOMA) provides the perfect complement to OMA [2]. In NOMA, it is feasible for users to access the channel nonorthogonally. OMA allows one user equipment (UE) per subcarrier, whereas NOMA permits many UEs to access the same subcarrier, which in turn produces multi-UE interference; it has been shown that this can be suppressed with appropriate interference cancellation (IC) and, as a consequence, this increases capacity [3]. All these factors certainly increase the potential of NOMA as a complementary technology to OMA in terms of accessing resources and supporting gigantic multiple-input multiple-output (gMIMO) connectivity [4], [5].

Since NOMA is a multiple-access scheme that achieved improved capacity for multiaccess channels, it has received much research attention as end-user computing power has grown [6]. [7] developed an optimum MIMO-NOMA power distribution scheme with user clustering to find a low-complexity solution to capacity restrictions. It has been demonstrated that in relation to energy efficiency (EE) and the quantity of consumers it can accommodate, the suggested method outperforms OMA and NOMA. [8] examined the effectiveness of NOMA and OMA in cellular networks using relevant theories and real-world problems. The Performance findings demonstrated that NOMA can make the performance for cell edge users much better than OMA can, which supports the use of NOMA in practice. [9] examined field tests on a NOMA microchip for a smartphone intended to improve spectral efficiency (SE). Terminals with a NOMA processor and single-user MIMO (SU-MIMO) were used for the trial. In comparison to using SU-MIMO in OMA mode, testing has shown that using NOMA with three terminals can increase the SE and data rate by up to 2.3 times.

Due to the potential benefits offered by NOMA and OMA, a hybrid multiple- access (HMA) scheme that focuses on these two principles to address future traffic demand was introduced in [10]. By adapting the hybrid NOMA/OMA scheme to the channel condition, [11] and [12] demonstrated the flexibility of the scheme and showed how it outperformed NOMA in terms of fairness when the strong user's channel gain was worse than a threshold set by the weak user's fixed target rate. In this context, a coalition-based user grouping approach was considered by the authors of [13] for resource allocation (RA) and the accommodation of a large number of Internet of Things (IoT) devices in an ultradense network. The scant research on RA in hybrid NOMA to date, including [11] and [13], has primarily examined channel allocation while attempting to optimize the throughput. However, only one antenna at the base station (BS) was considered in [12]. Given that modern BSs are typically equipped with multiple antennas [14], it is anticipated that maximum-ratio transmission (MRT) will demonstrate the usefulness of the hybrid multiple-access scheme. Overall, MIMO-HMA in conjunction with a small cell architecture is a strong candidate for next-generation network (NGN) success.

Perhaps the greatest challenge for gMIMO antenna realization under a particular specification or standard is EE. The application of the mechanism to perform a series of cluster-based HMA operations on the user distribution calls for enormous developments in design, connectivity services, and EE [15]. [16], [17], [18] developed joint optimization techniques for downlink (DL) and uplink (UL) transmissions to improve the way channels are assigned and the amount of power given to HMA users who use multiple utility functions, such as user fairness, capacity, EE, and quality of service (QoS) restrictions; however, they ignored the aspects of beamforming, channel state information (CSI), and mean squared error (MSE).

Multidirectional beamforming with each radio frequency (RF) chain interacts with power allocation and user pairing, as introduced in [19], in contrast to both analog beamforming and hybrid beamforming. Imperfect CSI was used in [20] and [21] to discuss how to build a beamforming method, which performs well but does not allow maximal-ratio transmission. [22] and [23] exploited spatial beamforming in MIMO to mitigate the intra- and interchannel/cluster interference brought by HMA but did not consider a heterogeneous network (HetNet) scenario in which both macrocells and small cells coexist. Using a lens antenna array, the method described in [24] aims to produce a narrow beam while maintaining fast channel tracking by underestimating the difficulties associated with nonorthogonal resource sharing. CSI, MIMO, and residual intersymbol interference (ISI) were not taken into account in [25], which proposed a hybrid beamforming approach for high-dimensional multiuser communication.

[26], [27] investigated a massive MIMO-HMA system by assuming perfect successive interference cancellation (SIC), which is almost impossible in practice. Issues with classical SIC were discussed in [28] and [29]. Two of these issues were accurate channel state information (CSI) estimation and hardware degradation. [30] identified various hardware and time-varying channel impairment concerns, and [31] described a general MIMO-NOMA method accounting for in-phase and quadrature-phase imbalance, as well as imperfect SIC where residual error is considered a complex Gaussian random variable. Unlike the above works, [32] introduced a joint design problem in which imperfect decoding and imperfect SIC were considered, but challenges of nonconvexity and nonlinearity remained. The NOMA scheme was studied by [33], utilizing an iterative interference cancellation (IIC) approach and turbo-coding single carrier (SC) MIMO. To mitigate residual ISI, the authors introduced iterative joint transmit/receive minimum MSE (MMSE) filtering but did not focus on EE optimization for hybrid resource access.

The requirement of joint iterative detection and decoding (JIDD) for soft IC was discussed in early works [34] and [35], in which the authors showed that JIDD could obtain better performance with soft IC using the MMSE technique than with maximum likelihood detection but ignored the

significance of digital modulation and detection. To estimate soft symbols, an MMSE-based joint iterative detection and decoding (JIDD) scheme was discussed that exploited gray-mapped quadrature amplitude modulation (QAM) and phase shift keying (PSK) in [36] and 8-PSK in [37]. In [38] and [39], the authors considered the MMSE model and investigated iterative IC (IIC) and channel estimation in which users do not have the choice of utilizing the same shared resource, either orthogonally or nonorthogonally depending on the channel condition. [40] converted the kZ code bits of a cooperative and recursive MUD-MMSE scheme into a series of Z complex symbols using 2^k-PSK modulation and examined how well this method performed over a quasistatic fading channel. To address RF impairment, [41] used M-QAM for multicarrier transmission, and [42] considered 64-QAM for wireless personal area networks, showing a reduction in phase noise (PN). The finite alphabet simplicity (FAS) algorithm and M-QAM with $M > 4$ were considered in [43], and it was claimed that this method achieved both a lower error rate and a lower complexity load.

The MSE-based transceiver was extensively studied in [5], [44], [45], [46], [47]. In [44] and [45], transceivers based on interference alignment (IA) were introduced to reduce the maximum MSE per stream. To fulfil the demand for high speed, EE, and SE in 5G, [46] presented a receiver design based on MMSE equalizers and a likelihood ratio test with an international telecommunication union (ITU) channel, where nonorthogonal resource utilization was disregarded. It was found that minimizing residual cross-symbol and cross-antenna interference results in improved efficiency. To improve convergence speed, an adaptive MMSE receiver with recursive least squares algorithms was proposed in [47] without QAM. Quadrature PSK (QPSK) and 16-QAM considered in [48] for an iterative multiuser detection (MUD) transceiver to decrease the block error rate (BLER). The authors of [49] stated that the best way to obtain the highest transmission rate is to use adaptive M-ary QAM and power adaptation. Additionally, [50] avoided considering time offsets and digital modulation to achieve a good transmission rate and added a new type of IC concept—different from SIC and parallel interference cancellation (PIC)—to the MMSE unit. They also provided a performance analysis in which the same shared resource was not used for either orthogonal and nonorthogonal transmission. However, none of the aforementioned studies combined the HMA approach and asynchronous transmission.

[51] and [52] examined asynchronous transmissions for HMA, where a technique called ''oversampling'' was used to ensure that symbols from different users arrived at different times. In terms of total throughput, the authors showed that asynchronous NOMA with a sufficiently long frame length outperformed synchronous NOMA. [53] proposed an IC technique for multiuser asynchronous NOMA transmissions. The paper's simulation and numerical findings demonstrate how significantly iterative IC can improve MA performance. The number of iterations relies on the transmission's detection strategy and modulation. A transceiver design for asynchronous NOMA transmissions without MRT or beamforming was reported in [54]. The asynchronism and MMSE scheme for HMA in the HetNet scenario, however, were not taken into account in the work mentioned above to suit consumer preferences.

A. MOTIVATION AND KEY CONTRIBUTIONS

The key limitations of the existing HMA architecture are as follows. First, none of the aforementioned studies considered the impact of digital modulation and detection on transceivers in the field of HMA. In fact, no attention has been given to HMA for the HetNet, where femto base stations (FBSs) coexist under the network coverage of a macro base station (MBS). Therefore, a gMIMO-HMA architecture integrating small cells and QAM is suggested as a consolidated solution to reduce the outage probability (OP) and increase the SE and EE, respectively. Second, prior works mostly focused on perfect SIC, which suffers from imperfect CSI, residual noise, and higher complexity. To overcome these disadvantages, a new IC scheme is suggested. Third, because users are typically dispersed geographically, the signals they send to the BS will arrive at the BS with time offsets due to the routes they took to get there. As a result, asynchronous transmissions for HMA are crucial, but they have not yet been thoroughly researched for the proposed iterative MMSE-IC-based transceiver.

Based on the discussion above, our significant contributions are summarized as follows:

- We propose a new HMA HetNet architecture that consists of small cells, MIMO, clustering, beamforming, M-QAM, and MMSE units where the existing users utilize the same shared resource either orthogonally or nonorthogonally.
- To overcome the major setbacks of traditional SIC (T-SIC), a novel transceiver that follows the principle of iterative MMSE-IC is proposed. The proposed transceiver accounts for both residual interference variance at the outer loop and PIC at the MMSE unit. An algorithm is also proposed for iterative mechanisms at the receiver. The closed-form formulas of normalized MSE and the best estimation coefficient for M-QAM in a noiseless environment are obtained according to Proposition 1, which is stated and demonstrated. In addition, for the problem formulation of the Outage Probability (OP) optimization subject to the given constraints, we first formulate the SINR expression under imperfect CSI.
- For a certain sum-rate of asynchronous transmission, a novel closed-form expression for the EE is derived, in which the Lagrange multiplier is considered to ensure that the optimization problem becomes a concave function of DL and UL time intervals. In addition, from the perspective of optimization, the necessary and sufficient conditions are given to maximize the EE.

TABLE 1. Notation of the main network parameters.

Symbol	Description
n_c	Number of UEs
$\{g_1, \cdots, g_{n_c}\}$	Channel gains related to different UEs where $g_1 > g_2, \cdots > g_{n_c}$
$\{m_1, \cdots, m_{n_c}\}$	A piece of information of a certain UE
N_o	AWGN experienced by a BS
$\{N_{o1}, \cdots, N_{on}\}$	AWGN experienced by the different UEs
v	Number of inputs of a multiplexer (MUX) or output of a DEMUX
r	Number of selection lines to send specific input lines to the output
$\{s_0 \cdots s_r\}$	A set of selection lines associated with a MUX or DEMUX
$\{Y_0 \cdots Y_v\}$	Outputs of QAM at R_x or Inputs of QAM at T_x
$\{X_0 \cdots X_v\}$	Outputs of SIC at R_x or inputs to PD-NOMA coding at T_x
$\{T_{x_0} \cdots T_{x_v}\}$	Outputs at the transmitter
$\{R_{x_0} \cdots R_{x_v}\}$	Inputs at the receiver
Z	Input of the MMSE-SIC optimal linear filter and IC
E	An enable input
A	Output of a MUX or input of a DEMUX
φ	Symbol frame length
δ_0	Symbol timing offset
t_0 and t_i	DL time interval and UL time interval
$\hat{g}_1$, $\hat{g}_2$ and $\hat{g}_{n_c}$	The quantized channel gains of g_1, g_2, and g_{n_c}
P_{max}	Maximum transmit power
p_1 and p_2	The transmit powers associated with UE 1 and UE 2
D and K	A $2L \times 2L$ matrix and channel matrix
$I_{2\varphi}$	A $2\varphi \times 2\varphi$ identity matrix
β^*	Optimal power co-efficient
y	Number of quantization levels
e	Number of bits in the quantization codeword
q	A scalar quantizer
λ_1	Lagrange multiplier
λ_2^* and λ_3^*	Lagrange multipliers in the optimum condition
$\mathbb{S}$	A set of t_i
R_{async}	Sum-rate in asynchronous mode
p_{DL} and p_{UL}	Power consumption during DL and UL wireless power transfer, respectively
ϕ_{EE}	EE in asynchronous mode
ω	Input signal vector
C_0	Leakage power
C_1 and C_2	Electronic and electric circuit powers
C_3 and C_4	Finite impulse response filter and beta coefficients

Proposition 2 is used to determine the two solutions of the quadratic equation, and proposition 3 is provided to find the optimal sum-rate expression for asynchronous mode.

- To the best of our knowledge, a comparative study with exemplary results for two modes within the framework of the proposed architecture is presented first in this paper. Furthermore, the use cases for hardware impairments in bridging the research gap to achieve standardization have been discussed but not yet investigated in the literature.

B. ORGANIZATION

The structure of this paper is as follows: Section II describes the proposed gMIMO-HMA architecture and various related technological enablers. We study the asynchronous mode of operation in Section III based on the discussion provided in Section II-D. Regarding the OP, SE, EE, and hardware impairment parameters, Section IV gives the simulation findings for four different approaches. The paper's conclusion is presented in Section V. Table 1 includes the parameter notation and descriptions that are considered in the paper.

II. PROPOSED ARCHITECTURE

Figure 1 and Figure 2 show an HMA beamformer-based architecture with cluster-based user associations for UL and DL transmissions. We explore the proposed architecture with beamforming, multiple antenna configurations, HMA, and MIMO to fulfil various user requirements such as QoS demands, massive UE connectivity, and interference.

A. HYBRID MULTIPLE ACCESS (HMA) SCHEME

The HMA scheme employs maximal-ratio transmission, which is based on the principles of both NOMA and OMA. It enables both of them to operate concurrently on the same shared resource and adaptively selects the best multiple-access approach based on the channel condition and user needs. To avoid the issue of having nearly the same channel gain within a cluster, cluster formation is determined by considering different SINR threshold values. A cluster consists of a number of users located in close proximity. There are two types of clusters based on strong users and weak users. It is noteworthy that there may be multiple clusters of each type. By taking into account a SINR threshold, a large number of users are divided into the two types.

A cluster can be created by taking into account users whose SINRs are higher than the threshold value, and a different cluster can be created by taking into account users whose SINRs are lower than the threshold value. One cluster (user's SINR>threshold) can be made even stronger by raising the threshold value; however, doing so would result in a sharp decline in the cluster's user population. However, in another cluster where the user SINRs are below the cutoff, the number of users will rise. The random selection of users from each cluster is followed by HMA clustering. The key idea is that the clusters may have different sizes; a small number of users might not be a part of any cluster, in which case OMA can be used. If a user receives network service from the closest FBS or MBS, the related BS, regardless of type, can adaptively switch between NOMA and OMA depending on whether it is paired or unpaired. The scheme offers greater SE and EE as well as more user connections because it is more dynamic than previous schemes. However, the signaling overhead of the HMA scheme increases in contrast to the scheme in which the shared resource is exploited either orthogonally or nonorthogonally.

B. KEY FUNCTIONAL MODULES

The deployment of femtocells, a type of small cell, in the macrocell infrastructure known as HetNets, which offers better coverage, higher SE, increased EE, and higher QoS, is necessary to meet the constantly growing customer demand for future mobile communication networks. For

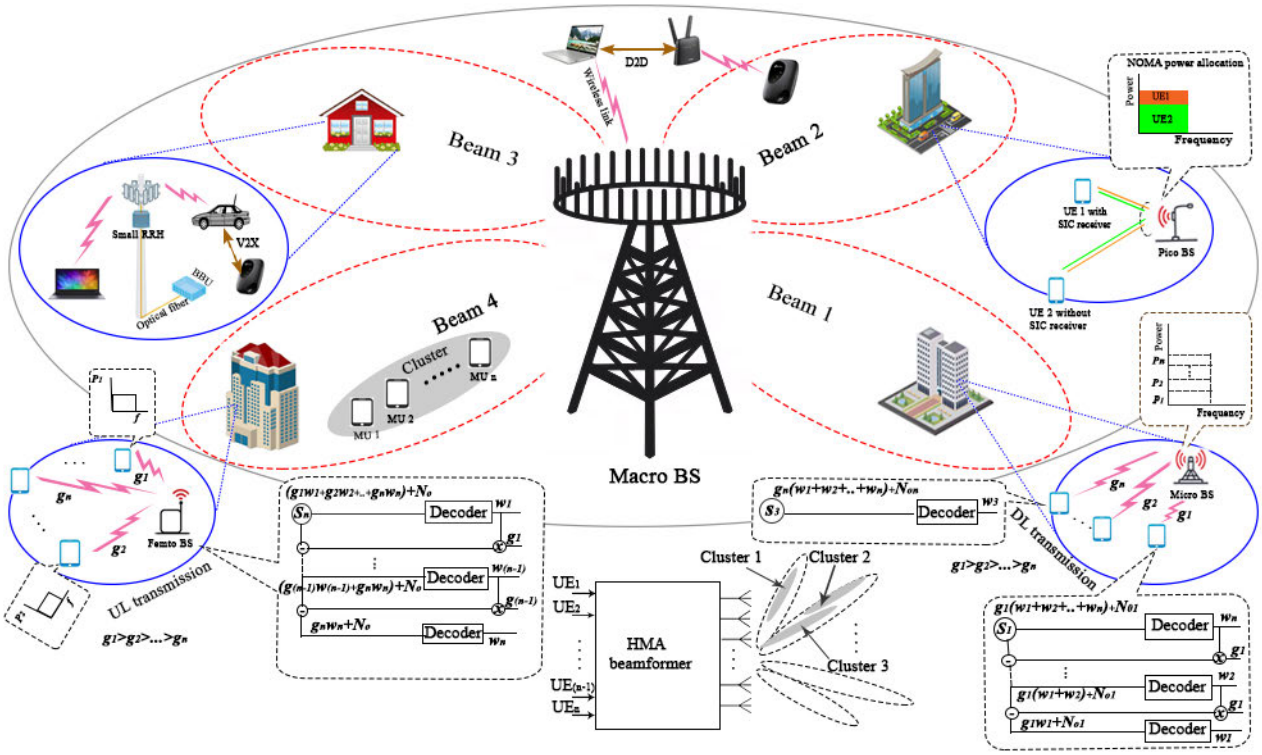


FIGURE 1. MIMO-HMA architecture: small-cells-integrated HetNets with a beamformer.

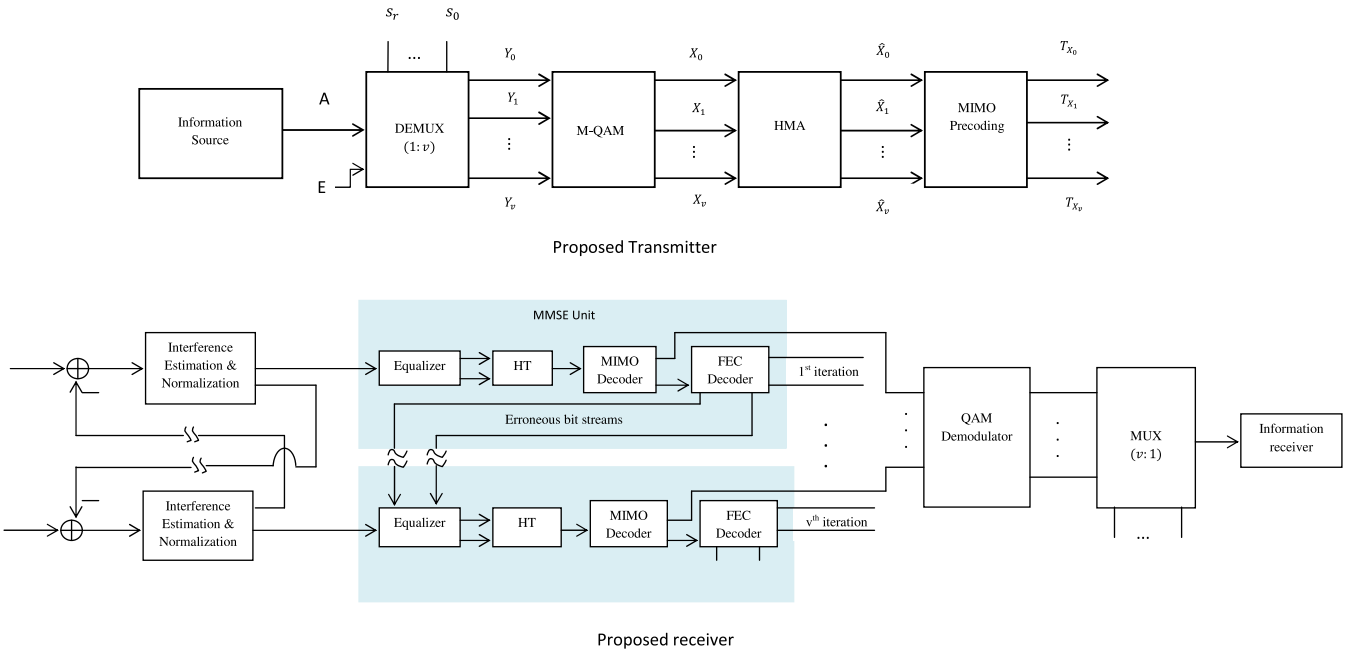


FIGURE 2. MIMO-HMA architecture: proposed transceiver architecture.

macrocells and femtocells, signal propagation and network coverage are very different. Small cells offer high-frequency coverage for approximately 100 meters, while macrocells offer low-frequency coverage over kilometers. Femtocells

can operate in licensed and unlicensed spectra, but they have a line of sight (LOS) problem.

Based on the most effective method of communication between source and destination, the mechanism of a cooperative cellular network architecture can be classified into three categories: i) DL transmission, where the source and UE communicate one-to-many; ii) UL transmission, where the source and UE communicate many-to-one; and iii) two-way transmission assisted architecture, where the source and UE communicate one-to-one. On the one hand, HMA is a challenging technique to use in UL communication due to the different channel conditions produced by different mobile users. Reduced limitations on signal synchronization at the transmitter and receiver sides, enhanced attainable sum rate, and decreased OP should be the goals of an effective UL-HMA network architecture framework. It has been demonstrated that better throughputs may be attained and that UEs experience less interference with increasing power differences between NOMA UEs in UL transmission.

In UL-HMA, several UEs emit signals in the form of waves across radio resource blocks with different regulated transmit powers allocated by the MA scheme. The signals that arrive at the small-cell BS are taken into account as the desired signal, even if they lead to unwanted interference with each other. At the small-cell base station (SBS), the SIC scheme can be used to decode a piece of information from a UE while preventing interference from other UEs. The signal of a UE in SIC operations is the received signal that is combined with the user's communication. Typically, a UE's channel gain can be used to estimate its signal intensity. Thus, in a general SIC setup, the message of a cluster's strongest UE is first decoded by canceling out the rest of the UE signals, and then the cluster's next strongest UE is decoded, and so on.

In DL-HMA, a small-cell BS transmits different signals across radio resource blocks[1] for the number of UE devices by using the multiuser transmission scheme of superposition coding. The desired signals, along with the severe interference produced due to other UE radiated signals, arrive at the recipient. To access the signal it requires, each UE with an SIC receiver decodes the signals in the composite signal. A UE with a significantly stronger channel gain is allocated a higher transmit power. Therefore, the message of the strongest UE in the cluster can be read, and the messages of weak UEs are removed one by one from the layered signal.

A synchronous mechanism is a type of full-duplex transmission method that uses the same clock pulse to let both the sender and the receiver talk at the same time. In contrast to the synchronous mechanism, which needs a common clock pulse from the outside, the asynchronous mechanism is a half-duplex type of communication in which the transmitter and receiver have their own internal clocks. In synchronous transfer, data are sent in frames. On the other hand, asynchronous transfer sends one byte of data at a time. The interval between two succeeding transmissions is constant in synchronous operation. Asynchronous operation results in a random delay between two subsequent messages. With the synchronous approach, much information can be transferred quickly. In comparison, asynchronous transfer has a slow data transfer rate.

Even though asynchronous transfer has many drawbacks, such as timing error, bit reversal due to noise, and additional overhead due to the inclusion of extra bits to start and stop the transmission, it has the following benefits over synchronous transfer: the receiver and transmitter do not need to be synchronized, users can transmit signals at different bit rates, and after one data byte transmission is finished, a user can start another one without waiting to obtain a response from the server.

C. HMA BEAMFORMER WITH A MULTIANTENNA CONFIGURATION

Beamforming is a subset of gMIMO. In general, beamforming uses multiple antennas to control the direction of a wave front by appropriately weighting the magnitude and phase of individual antenna signals in an array of multiple antennas. This gMIMO-HMA system model is based on the cochannel deployment of a macrocell with massive antenna arrays and femtocells with multiple antennas each. Despite possible transceiver hardware impairments, it has been shown that massive antenna arrays at BSs can significantly enhance spatial resolution and array gain, which in turn improve SE. Additionally, massive antenna arrays cause channel decorrelation between users, which can result in a better signal gain with nominal interuser interference. The use of HMA benefits from this kind of channel characteristic. When beamforming is used with gMIMO transmissions, it opens up new possibilities that could greatly increase the total network throughput. To completely mitigate interbeam interference, it is sufficient to have more antennas at the transmitter or at least the same antenna counts at the receiver. On the other hand, HMA forbids cobeam interference. Although underprivileged users of a cluster typically encounter a significant quantity of strong cross-cluster interference and a minimal amount of cocluster interference, this typically causes the signal-to-interference plus noise ratio (SINR) to decrease.

The combined scenario of gMIMO-HMA beamforming can allow virtually interference-free communications regardless of the spatial distribution. UEs distributed over a large spatial domain can benefit from beamforming for interference-free communications, and UEs located in a small spatial domain can benefit from using HMA IIC to suppress interference. The use of numerous antennas and spatial multiplexing at the transceivers aid in achieving capacity gain while mitigating the significant route loss of high- frequency signals. Although gMIMO provides many advantages, it makes fully digital beamforming costly as each

[1]Radio resource blocks are described as a series of adjacent subcarriers in the frequency domain.

antenna needs an individual RF chain for signal processing. This problem can be fixed in part by combining digital and analog precoding in the design [25]. This not only improves EE but also decreases the number of RF chains that are needed. With the beam design, it can be very helpful to put a few IoT devices close together to form a cluster. However, clustering is achievable with NOMA with fewer RF chains, whereas with OMA, each device requires its own RF chain, which increases the cost and size of the design.

D. PROPOSED TRANSCEIVER ARCHITECTURE

Illustrations of gMIMO-HMA transmitter[2] and receiver [55] architectures are also shown in Figure 2, where v, r, $\{s_0 \cdots s_r\}$, $\{Y_0 \cdots Y_v\}$, $\{X_0 \cdots X_v\}$, E, A are defined as in Table 1. On the transmitter side, the QAM and HMA are the two crucial signal processing stages for optimum transmitter performance. These stages need to work in harmony with the IC for HMA and the digital demodulator stage at the receiver side to ensure a reliable and high-capacity transmission. A precoder at the transmitter is used to transmit a coded data stream to the receiver to acquire prior knowledge of the channel. Baseband signals are first routed through a low-dimensional digital precoder to ensure performance and then through an analog precoder to improve EE and reduce hardware complexity.

Currently, the NGN model uses QAM to modulate the amplitudes of two orthogonal carrier waves using amplitude-shift keying (ASK) to convey two digital bit streams simultaneously. The exploitation of QAM helps to achieve high SE. Depending upon the demand, multilevel detection may take place, where the size and center point of the constellation map for each level may need to be changed.

IMMSE is the CSI technique used here to estimate the known statistical characteristic of a channel. IMMSE both reduces residual error and shortens the training sequence. During the training process, a known signal (known as a pilot) is sent from the transmitting side, and the receiving side uses the combined knowledge of the transmitted and received signals to estimate the channel matrix $K_{\varkappa}^{\chi}$. This a priori information can be used to reduce estimation errors if the channel and noise distribution are known. To avoid MSE degradation, accurate knowledge of the channel correlation matrix and noise correlation matrix should be available. To equalize the received signal, the receiver makes use of the estimated channel characteristics. Prior to decoding the transmitted signal, the distortion brought by the wireless channel is eliminated by the equalization process. At each iteration in the MMSE unit, the equalizer coefficient is computed to determine each channel. One codeword is made up of L_p parallel filters and L_m modulated symbols for a codeword length of $L_p L_m \varrho$ bits. For all antennas, coded bitmapping is performed on an M-QAM constellation with a cardinality of 2^{ϱ}. Each channel uses a portion of the source data, consisting of $\log_2 M$ bits, which are mapped to an M-ary symbol in the M-QAM modulation scheme and subsequently released from the RF chain to transmit information. To mitigate the impact of burst error, deinterleaving and interleaving are performed at the MIMO decoder stage. If erroneous bit streams are detected, the forward error correction (FEC) detector's role is to continue storing a priori statistics such as the mean and variance. The hypothesis tests (HTs) are generated based on a priori statistics and soft demapping from the equalizer. The output of the HT is important for the interleaver and demodulator.

We extend the investigation of the MMSE unit, presented in [55], by considering its application for the proposed transceiver as a part of the design. The considerations of the MIMO decoder and multistage iterations improve the effectiveness of the MMSE unit compared to that in [55]. In comparison to [55], derivations, propositions, and an iterative MMSE-IC algorithm are added to the analytical part of the MMSE unit. Aside from the transceiver, the signal processing at the receiver goes through the following steps sequentially: first, interference estimation and normalization are used to account for residual errors caused by imperfect SIC; next, the MMSE unit is used to correct the errors; and finally, QAM demodulation and multiplexing are used to reconstruct the signal before it is received at the final stage.

Let the modulating signal vector be given by

$$s = K_{\varkappa}^{\chi} F \omega + n_0, \tag{1}$$

where $K_{\varkappa}^{\chi}, F, \omega, n_0$ denote the channel matrix, precoding matrix, transmitted symbol vector, and independent identically distributed (i.i.d.) additive white Gaussian noise (AWGN) with zero mean and σ_1^2 variance. In a HetNet environment, a user is said to receive network service from an MBS if $\chi = m$, while a user is said to be receiving network service from an FBS if $\chi = f$. In addition to the HMA strategy, if the base station chooses OMA, then $\varkappa$ indicates OMA, and if it chooses NOMA, then $\varkappa$ indicates NOMA. They can further be defined as follows:

$$E\left[FF^H\right] = I, \tag{2}$$

$$E\left[\omega\omega^H\right] = I_{n_t}, \tag{3}$$

$$E\left[n_0 n_0^H\right] = \sigma_1^2 I_{n_r}. \tag{4}$$

$$K_{OMA}^{\chi} = \begin{bmatrix} g_1\sqrt{p} & 0 & 0 & 0 & 0 \\ 0 & g_2\sqrt{p} & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & g_{n_c-1}\sqrt{p} & 0 \\ 0 & 0 & 0 & 0 & g_{n_c}\sqrt{p} \end{bmatrix}, \tag{5}$$

[2] $v = 2^r$, $r = \log_2 v$, $Y_0 = E\bar{s}_r \cdots \bar{s}_1 \bar{s}_0 A$, $Y_1 = E\bar{s}_r \cdots \bar{s}_1 s_0 A, \cdots$, $Y_n = E s_r \cdots s_1 s_0 A$ for $E = 1.A = \bar{s}_r \cdots \bar{s}_1 \bar{s}_0 Y_0 + \bar{s}_r \cdots \bar{s}_1 s_0 Y_1 + \cdots + s_r \cdots s_1 s_0 Y_1$
$A = \bar{s}_r \cdots \bar{s}_1 \bar{s}_0 Y_0 + \bar{s}_r \cdots \bar{s}_1 s_0 Y_1 + \cdots + s_r \cdots s_1 s_0 Y_1$.

$$K^{\chi}_{NOMA} = \begin{bmatrix} g_1\sqrt{p_1} & 0 & 0 & 0 & 0 \\ 0 & g_2\sqrt{p_2} & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & g_{n_c-1}\sqrt{p_{n_c-1}} & 0 \\ 0 & 0 & 0 & 0 & g_{n_c}\sqrt{p_{n_c}} \end{bmatrix}, \tag{6}$$

where $\{g_1, g_2, \cdots, g_{n_c}\}$ denote channel gain; I, I_{n_t}, and I_{n_r} denote identity matrices. If ξ_i and $\varrho_{j,i}$ denote the fading coefficient and average received power per antenna element, respectively, the power associated with the j^{th} IoT device is given by

$$p_j = \sum_i \xi_i \varrho_{j,i} = \sum_i \varrho_{j,i}\xi_i, \tag{7}$$

and (7) may also be redefined as follows:

$$\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n_c-1} \\ p_{n_c} \end{bmatrix} = \begin{bmatrix} \varrho_{1,1} & \varrho_{1,2} & \cdots & \varrho_{1,n_r-1} & \varrho_{1,n_r} \\ \varrho_{2,1} & \varrho_{2,2} & \cdots & \varrho_{2,n_r-1} & \varrho_{2,n_r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \varrho_{n_c-1,1} & \varrho_{n_c-1,2} & \cdots & \varrho_{n_c-1,n_r-1} & \varrho_{n_c-1,n_r} \\ \varrho_{n_c,1} & \varrho_{n_c,2} & \cdots & \varrho_{n_c,n_r-1} & \varrho_{n_c,n_r} \end{bmatrix} \times \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{n_r-1} \\ \xi_{n_r} \end{bmatrix}. \tag{8}$$

The residual error variance can be expressed as

$$\tilde{K} = E\left[(K^{\chi}_{\varkappa} - \hat{K}^{\chi}_{\varkappa})^2\right], \tag{9}$$

where $\hat{K}^{\chi}_{\varkappa}$ denotes the estimation of interference. If $\hat{Z}$ denotes an $n_t \times n_r$ detector matrix at an MMSE unit, then it can be expressed by incorporating imperfect CSI as

$$\hat{Z}_{MMSE} = E\left[(\tilde{K}\tilde{K}^H + I_{n_r})^{-1}\tilde{K}\right], \tag{10}$$

where H is Hermitian Transpose. After linear detection at the MMSE unit, (1) can further be expressed as

$$S = \hat{Z}^H_{MMSE} K^{\chi}_{\varkappa} F\omega + \hat{Z}^H_{MMSE} n_0. \tag{11}$$

Furthermore, the estimated output at the MMSE unit can be represented by

$$\tilde{\omega} = \left[[K^{\chi}_{\varkappa}]^H W^H \left(S - K^{\chi}_{\varkappa} F\hat{\omega}\right) + \mathrm{diag}(\delta_2)\hat{\omega}\right] \times \mathrm{diag}(\delta_1), \tag{12}$$

where δ_1 and δ_2 denote equalizer coefficients, $E[(s-\mu)(s-\mu)^H]$ and μ indicate the covariance matrix and mean of the signal after PIC, respectively, and W stands for the equalization matrix as follows:

$$W = \left[(\tilde{K} - E[(s-\mu)(s-\mu)^H])K^{\chi}_{\varkappa}[K^{\chi}_{\varkappa}]^H + \sigma_1^2 I_{n_r}\right]^{-1}. \tag{13}$$

Algorithm 1 Iterative MMSE-IC Algorithm

```
Initialization: F = I_{n_t}, Q = I_{n_r}, iteration index (i) = 1, maximum iteration = n_I
while i < n_I do
    Compute: E[(K^χ_ϰ − K̂^χ_ϰ)^2]; \\ variance of residual interference
    Compute: E[(s − μ)(s − μ)^H]; \\ variance belongs to MMSE unit
    Determine SINR by (16)
    Normalize F and Q to satisfy the optimization constraints
    Check the subset of erroneous bit streams: Ψ_e
    if |Ψ_e| = 0 then
        break
    else
        Perform PIC
        {F, β*} ← arg min E(||ω − β*^{-1} ω̃||^2)
        i = i + 1
    end if
end while
```

Proposition 1: For PIC efficiency

$$\eta_{IC} = [1 - (\text{normalized interference})],$$

the normalized MSE for a noiseless environment is

$$\epsilon = E\left[\frac{|K^{\chi}_{\varkappa}F - \hat{K}^{\chi}_{\varkappa}\hat{F}|^2}{|K^{\chi}_{\varkappa}FM|^2}\right], \tag{14}$$

and the optimal estimation coefficient for M-QAM is

$$\zeta^* = \frac{E\left[Re\{\hat{\omega}M\}\right]}{E\left[|\hat{\omega}|^2\right]\left[1 + M^2\epsilon\right]}. \tag{15}$$

Proof: APPENDIX A shows the proof. □

For the iterative mechanism at the receiver, an iterative MMSE-IC algorithm is proposed, as shown below.

E. PROBLEM FORMULATION

The SINR of the j^{th} IoT device can be expressed by

$$\gamma_j = \frac{\hat{Z}^H_{MMSE,j} K^{\chi}_{\varkappa,j} F_j}{\sum_{\substack{a=1,\\ a\neq j}}^{n_c} \hat{Z}^H_{MMSE,a} K^{\chi}_{\varkappa,a} F_a + \sigma_1^2 \hat{Z}^H_{MMSE,a}}. \tag{16}$$

Let n_c be the number of devices located in close proximity forming a cluster, if the total number of clusters belonging to narrow beam coverage is m, then the SINR maximization problem can be given by

$$\max_{P,F,K^{\chi}_{\varkappa}} \sum_{i=1}^{m}\sum_{j=1}^{n_c} \gamma_{i,j}, \tag{17}$$

Algorithm 2 Threshold Optimization Algorithm

Initialization: $\epsilon = 0.01$, $\xi_{const} = 0.2$, $\gamma_{const} = 0.1$, $\mathbb{P}_{min}^{out} = 0.5$
while do $\xi_i \leq 0.6$
  set $\gamma_{th,i} = \gamma_{const}$
  while $\gamma_{th} \leq \frac{\xi_{const}}{1-\xi_{const}}$ **do**
    Compute ϵ by (14)
    if $\epsilon \leq \epsilon_{const}$ **then**
      Determine $\mathbb{P}_j^{out}$ by (20)
      if $\mathbb{P}_j^{out} \leq \mathbb{P}_{min}^{out}$ **then**
        $\xi_i^* = \xi_{const}$, $\gamma_{th,i}^* = \gamma_{th,i}$
        $\mathbb{P}_{min}^{out} = \mathbb{P}_j^{out}$
      end if
    end if
    $\gamma_{th,i} = \gamma_{th,i} + \gamma_{const}$
  end while
  $\xi_i = \xi_i + \xi_{const}$
end while
Return ξ_i^*, γ_{th}^*

$$\text{subject to: } tr(FF^H) \leq P_{max}, \tag{18}$$

$$tr(Q(K_\varkappa^\chi FF^H[K_\varkappa^\chi]^H + \sigma_1^2 I_{n_r})Q^H) \leq \xi_i \varrho_{j,i}. \tag{19}$$

We consider that an outage event[3] occurs in the HMA scheme if any user within a cluster is unable to decode another user belonging to the same cluster. The OP of the j^{th} IoT device in a cluster can be expressed by

$$\mathbb{P}_j^{out} = 1 - Pr\left\{\bigcap_{i=1}^{j} \gamma_j^i \geq \gamma_{th,j}\right\}, 1 \leq j \leq n_c, \tag{20}$$

where $\gamma_{th,j}$ indicates the threshold SINR to successfully decode the j^{th} user. For a given $\xi_i \varrho_{j,i}$ associated with the i^{th} antenna element and the j^{th} user, the OP minimization problem can be written as

$$[\xi_i^*, \gamma_{th,j}^*] = \arg\min_{\xi_i, \gamma_{th,j}} \mathbb{P}_j^{out}, \tag{21}$$

$$\text{subject to: } \gamma_{th,i} \geq 0, \tag{22}$$

$$\xi_{i+1}^* > \xi_i^* \gamma_{th,j}^*, \tag{23}$$

$$\epsilon \leq \epsilon_{const}. \tag{24}$$

To achieve the best performance from (21), we apply the threshold optimization algorithm, as shown in Algorithm 2.

III. ASYNCHRONOUS MODE OF OPERATION

By following [56], the achievable sum-rate in an asynchronous mode of operation can be expressed as

$$R_{async} = \frac{1}{\varphi + \delta_0} \log_2 \begin{vmatrix} I_{2\varphi} + K_\varkappa^\chi F \hat{Z}_{MMSE} \\ QF^H[K_\varkappa^\chi]^H \hat{Z}_{MMSE}^H \end{vmatrix}, \tag{25}$$

where φ represents the symbol frame length, δ_0 ($0 \leq \delta_0 \leq 1$) denotes the timing offsets as users' signals drift apart, $I_{2\varphi} \in \mathbb{C}^{2\varphi \times 2\varphi}$ indicates an identity matrix, $(.)^H$ denotes the Hermitian transpose, and Q is defined by (26).

$$Q = \begin{bmatrix} 1 & 1-\delta_0 & 0 & \dots & \dots & 0 \\ 1-\delta_0 & 1 & \delta_0 & 0 & \dots & 0 \\ 0 & \delta_0 & 1 & 1-\delta_0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \delta_0 & 1 & 1-\delta_0 \\ 0 & \dots & \dots & 0 & 1-\delta_0 & 1 \end{bmatrix}. \tag{26}$$

It is worth noting that (25) holds true for a finite value of φ. However, we can re-express (25) for $\varphi \to \infty$ as

$$R_{async} = \frac{1}{\varphi(1 + \frac{\delta_0}{\varphi})} \log_2 \begin{vmatrix} I_{2\varphi} + K_\varkappa^\chi F \hat{Z}_{MMSE} \\ QF^H[K_\varkappa^\chi]^H \hat{Z}_{MMSE}^H \end{vmatrix}$$

$$\cong \frac{1}{\varphi} \log_2 \Gamma, \tag{27}$$

where Γ is defined as

$$\Gamma = \begin{vmatrix} I_{2\varphi} + \beta^* K_\varkappa^\chi F \hat{Z}_{MMSE} \\ QF^H[K_\varkappa^\chi]^H \hat{Z}_{MMSE}^H \end{vmatrix}. \tag{28}$$

Here, (27) is similar to the expression associated with synchronous mode.[4] Most importantly, considering δ_0 as a whole makes designing a practical system even more feasible. Hence (27) becomes

$$R_{async} = \frac{1}{\varphi + 1} \log_2 \Gamma. \tag{29}$$

Therefore, the minimum rate can be maximized to provide equality among the UEs, and the following can be obtained from (25) with the inclusion of the optimal power coefficient:

$$R_{async}^* = \frac{1}{\varphi + \delta_0} \log_2 \Gamma, \tag{30}$$

where $\hat{g}_{min} = \min\{\hat{g}_1, \hat{g}_2, \cdots, \hat{g}_{n_c}\}$ in which we quantize[5] the channel gain g_1, g_2, and g_{n_c} as $\hat{g}_1 = q|g_1|^2$, $\hat{g}_2 = q|g_2|^2$, and $\hat{g}_{n_c} = q|g_{n_c}|^2$, respectively, applying a scalar quantizer denoted by q; β^* is defined as

$$\beta^* = \frac{2\hat{g}_{min}}{\left[\sqrt{(\hat{g}_1 + \hat{g}_2 + \cdots + \hat{g}_{n_c})^2 + 4p\hat{g}_1\hat{g}_2\hat{g}_{min}} + \hat{g}_1 + \hat{g}_2 + \cdots + \hat{g}_{n_c}\right]}. \tag{31}$$

By following this property,[6] the term Γ from (30) can be further reexpressed as

$$\Gamma = \left|\beta^* K_\varkappa^\chi F \hat{Z}_{MMSE}[K_\varkappa^\chi]^H F^H \hat{Z}_{MMSE}^H\right| \times \left|(\beta^* K_\varkappa^\chi F \hat{Z}_{MMSE}[K_\varkappa^\chi]^H F^H \hat{Z}_{MMSE}^H)^{-1} + Q\right|$$

[3] An outage refers to an underprivileged user.

[4] The achievable sum-rate for synchronous operation is given by $R_{sync} = \frac{1}{\varphi} \log_2 \left|I_{2\varphi} + K_\varkappa^\chi F \hat{Z}_{MMSE} Q F^H [K_\varkappa^\chi]^H\right|$ as the peaks of all the NOMA users' signals are perfectly aligned.

[5] Quantization is performed by considering $y = 2^e$ levels such as $\{q_0, q_1, \cdots, q_x\}$, where e denotes the number of bits in the quantization codeword.

[6] For any invertible $m \times m$ matrix, Z : $|Z + AB| = det(Z)det(I + AZ^{-1}B)$, where the matrix is composed of four submatrices Z, A, B, and I of dimension $m \times m$, $m \times n$, $n \times m$ and $n \times n$, respectively.

$$= (\beta^* p_1 g_1^2)^{\varphi} (\beta^* p_2 g_2^2)^{\varphi} \cdots (\beta^* p_{n_c} g_{n_c}^2)^{\varphi} |\Lambda|. \quad (32)$$

where Λ is defined as

$$\Lambda = (\beta^* K_{\varkappa}^{\chi} F \hat{Z}_{MMSE} F^H [K_{\varkappa}^{\chi}]^H \hat{Z}_{MMSE}^H)^{-1} + Q \quad (33)$$

Proposition 2: If $D_{u,v}$ denotes the submatrix of a matrix $((\beta^* K_{\varkappa}^{\chi} F \hat{Z}_{MMSE} F^H [K_{\varkappa}^{\chi}]^H \hat{Z}_{MMSE}^H)^{-1} + Q)$, the two solutions of the associated quadratic equation are given in (35) and (36), as shown at the bottom of the next page.

Proof: APPENDIX B shows the proof. □

Proposition 3: For (30), (32) and the $2\varphi \times 2\varphi$ matrix Λ given in (33), R_{async}^* is given by

$$R_{async}^* = \frac{1}{\varphi + \delta_0} \left\{ \varphi \log_2(a_1 a_2) + \log_2 \left[(k_1 + k_2)^2 - k_1 k_2 (1 + a_1^{-1}) \right] \right\}. \quad (38)$$

Proof: APPENDIX C shows the proof. □

Therefore, EE analysis can be expressed as shown below in an asynchronous mode of operation (37), as shown at the bottom of the next page, where $t_i = [t_1, t_2, \cdots, t_N]$ and p_{DL} and p_{UL} represent power usage during DL and UL wireless power transfer, respectively. Thus, the following is an expression of the optimization problem:

$$\max_{t_0, t_i} \phi_{EE}, \quad (39)$$

$$\text{subject to: } \sum_{i=1}^{\varphi} t_i < 1 - t_0, \quad (40)$$

$$R_i \geq R_{min}^i, \quad (41)$$

where R_i is defined as

$$R_i = \sum_{i=1}^{\varphi} \frac{1}{\varphi + \delta_0} \left\{ \varphi \log_2(a_1 a_2) + \log_2 \left[(k_1 + k_2)^2 - k_1 k_2 (1 + a_1^{-1}) \right] \right\}. \quad (42)$$

The Lagrangian of (42) can be given by [(30) in 83]

$$\mathcal{L}(t_0, t_i) = R_i - \lambda_1 \left[p_{DL}(t_0 + \sum_{i=1}^{\varphi - 1} t_i) + p_{UL} \sum_{i=1}^{\varphi} t_i \right],$$

where λ_1 denotes the nonnegative Lagrange multiplier. According to (42), R_i is a concave function of t_i, and the optimization problem (39) can further extend to a concave function of DL and UL temporal periods, i.e., t_0 and t_i, as follows:

$$\min_{t_i \in \mathbb{S}, \lambda_1} \left\{ \max_{t_0, t_i} \mathcal{L}(t_0, t_i) \right\},$$
$$\text{subject to: (40) and (41)}, \quad (43)$$

where $\mathbb{S}$ is a set of t_i that satisfy (40) and (41).

By following the inequality constraints, the Karush–Kuhn–Tucker (KKT) method can provide the necessary and sufficient conditions as follows

$$\frac{\partial \zeta}{\partial t_o} = 0 \text{ at } t_0 = t_0^*, \ \frac{\partial \zeta}{\partial t_i} = 0 \text{ at } t_i = t_i^*, \quad (44)$$

TABLE 2. Simulation configuration.

Parameter	Value
Radius of macrocell	3500 m
Interference protection radius of macrocell	500 m
Interference protection radius of small cell	200 m
Beam gain	14 dB
Ambient temperature	250 K
Channel coherence	64 blocks
Number of subcarriers	80
Channel width	20 MHz
Noise margin	4 dB
Data rate across multiple channels	10 Gbps
Wavelength	0.15 m
T_x power of target user	0 dB
T_x power of rest users	0-8 dB
g_1	$\sim \exp(0.4)$
g_2	$\sim \exp(0.9)$
Bit error rate	10^{-6}
Overloading factor	170%
y	16
Time duration of a symbol	1 μs
L	20
t	0.4

$$\lambda_2^* (R_i - R_{min}^i) = 0, \quad \forall i = \mathbb{S}, \quad (45)$$

$$\lambda_3^* \left(1 - \sum_{i=0}^{L} t_i \right) = 0, \quad (46)$$

where $\zeta = \mathcal{L}(t_0, t_i) - \sum_{i=1}^{L} \lambda_2^i (R_i - R_{min}^i) - \lambda_3 (1 - \sum_{i=0}^{L} t_i)$.

IV. SIMULATION RESULTS

In this part of the work, a single macrocell-based network architecture for gigantic MIMO-HMA communication that uses the MMSE-IC scheme is considered. Table 2 gives the details of the simulation configuration. The small-scale fading of the channels is considered Rayleigh fading for predicting the OP of the user's SINR for a given threshold of 0.2. Because UEs are dispersed at random by a Poisson point process throughout the simulation region, it is possible to compile an average over fast fading with a noise variance of 0.01 for each UE. The beamforming vector is developed based on the consideration of imperfect CSI, as the transmitters in DL do not carry the CSI due to the uncertainty of the channel vector.

A point worth noting is that the need for CSI increases with multiple antennas, and the process of CSI estimation and acquisition may deteriorate reliability. To preserve QoS for users, the error covariance matrix threshold is set to 0.006. The findings from numerous random channels are averaged to provide the performance evaluation statistics below. We compare the asynchronous mode (async mode) and the synchronous mode (sync mode) of signal processing for the reception of signals in the HMA. In sync mode, the peaks of all the HMA users' signals are perfectly aligned. In async mode, the peaks of the HMA users' signals drift apart, which creates timing offsets between the signals and can significantly deteriorate the received signal if it is not appropriately processed.

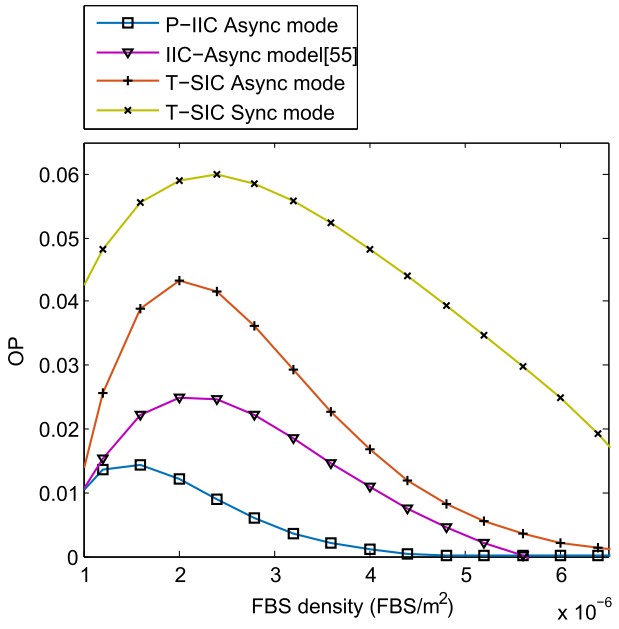


FIGURE 3. OP versus FBS density per square meter.

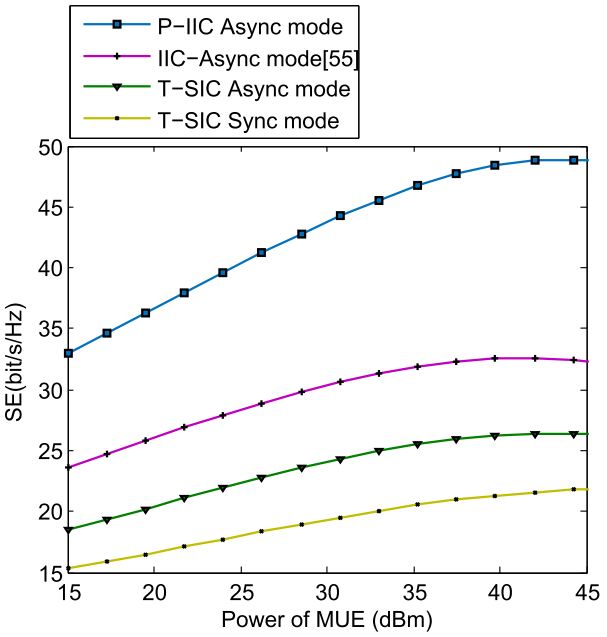


FIGURE 4. SE versus UE transmit power.

Figure 3 depicts the OP performance of the user's SINR versus FBS density for the proposed IIC (P-IIC) async, IIC async [55], T-SIC async, and T-SIC sync schemes. Notably, the OPs of all the schemes are high when the FBS density is between 1 and 2.5, except P-IIC async. The figure shows that when the correct signal processing is performed, OP decreases as FBS density increases, and the response is more favorable for asynchronism than for synchronism. As the FBS density increases, the OPs of all the schemes related to the async mode rapidly decline. Although the T-SIC async strategy surpasses the T-SIC sync method and IIC-async [55] once again outperforms the T-SIC async strategy, the P-IIC async strategy exhibits a substantial performance gap compared to the other three strategies, validating the viability of the suggested strategy. The key cause of the reduced performance in the other two T-SIC-based schemes is increased residual interference in the different stages of the SIC mechanism. On the other hand, the P-IIC async scheme can effectively reduce residual interference, which in turn produces better OP than other schemes, even

$$(\beta^* K_x^{\chi} F \hat{Z}_{MMSE} [K_x^{\chi}]^H F^H \hat{Z}_{MMSE}^H)^{-1} + Q$$
$$= \begin{bmatrix} 1+(\beta^* p_1 g_1^2)^{-1} & 1-\delta_0 & 0 & \dots & \dots & 0 \\ 1-\delta_0 & 1+(\beta^* p_2 g_2^2)^{-1} & \delta_0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 1-\delta_0 & 1+(\beta^* p_2 g_2^2)^{-1} & \delta_0 & 0 \\ 0 & \dots & 0 & \delta_0 & 1+(\beta^* p_{n_c-1} g_{n_c-1}^2)^{-1} & 1-\delta_0 \\ 0 & \dots & \dots & 0 & 1-\delta_0 & 1+(\beta^* p_{n_c} g_{n_c}^2)^{-1} \end{bmatrix}. \quad (34)$$

$$k_1 = \frac{a_1^{-1} + a_2^{-1} + a_1^{-1}a_2^{-1} + \delta_0(2-\delta_0) + \sqrt{(a_1^{-1} + a_2^{-1} + a_1^{-1}a_2^{-1} + \delta_0(2-\delta_0))^2 - 4\delta_0^2(1+a_2^{-1})}}{2}, \quad (35)$$

$$k_2 = \frac{a_1^{-1} + a_2^{-1} + a_1^{-1}a_2^{-1} + \delta_0(2-\delta_0) - \sqrt{(a_1^{-1} + a_2^{-1} + a_1^{-1}a_2^{-1} + \delta_0(2-\delta_0))^2 - 4\delta_0^2(1+a_2^{-1})}}{2}. \quad (36)$$

$$\phi_{EE} = \frac{\sum_{i=1}^{\varphi} \frac{1}{\varphi+\delta_0} \left\{ \varphi \log_2(a_1 a_2) + \log_2 \left[(k_1+k_2)^2 - k_1 k_2 (1+a_1^{-1}) \right] \right\}}{p_{DL} \left(t_0 + \sum_{i=1}^{\varphi-1} t_i \right) + p_{UL} \sum_{i=1}^{\varphi} t_i}. \quad (37)$$

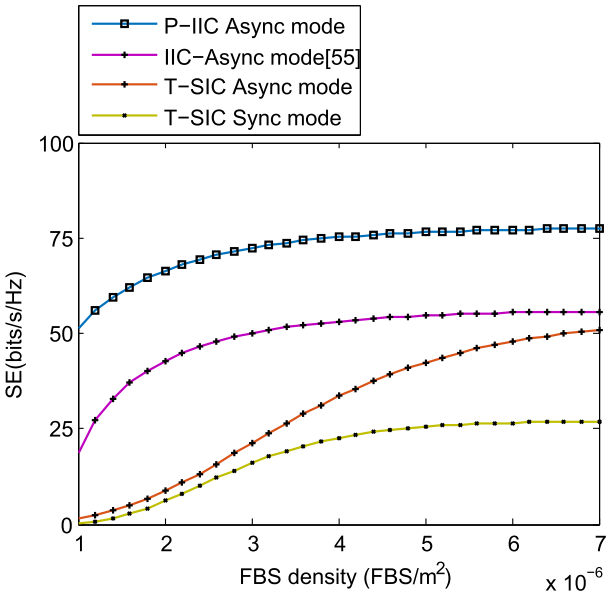


FIGURE 5. SE versus FBS density per square meter.

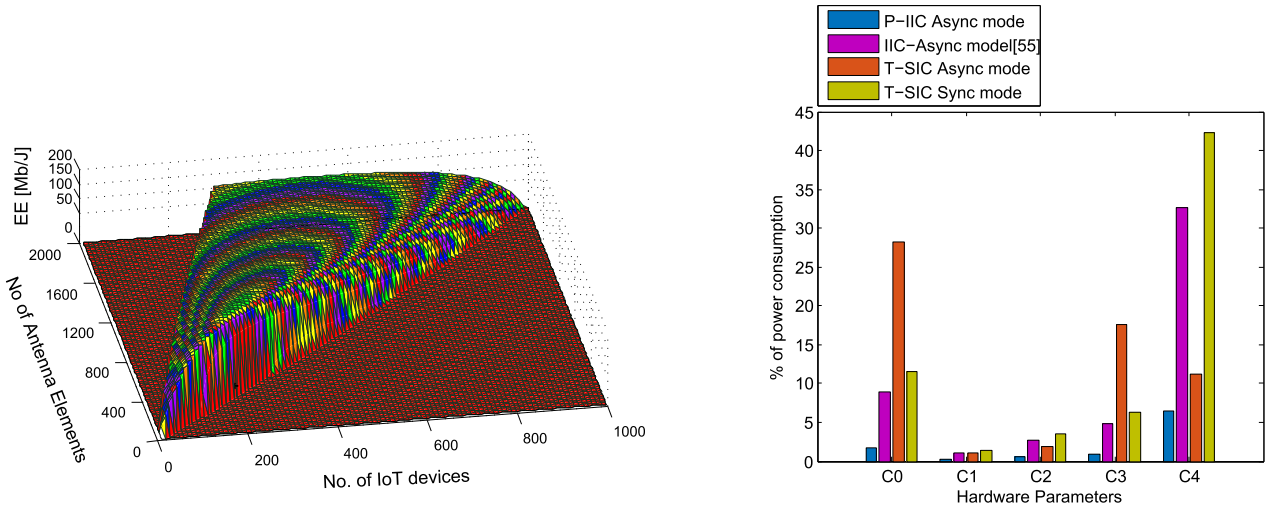
IIC async [55]. The proposed scheme iteratively decreases residual interference at the receiver by considering both residual interference variance at the outer loop and PIC at the MMSE unit. A high data rate is also made possible by the transceiver design's incorporation of M-QAM, which also yields a significant improvement in the proposed scheme's overall performance.

Figure 4 shows the SE performance versus UE transmit power (Multi-User Equipment - MUE in dBm) for the P-IIC async, IIC async [55], T-SIC async, and T-SIC sync schemes. While the performance of the scheme in [55] is only marginally improved when SISO is employed, both the P-IIC and T-SIC schemes perform noticeably better when asynchronism is taken into consideration than when it is ignored. For the async mode, the performance increases as the UE transmission power increases up to 40 dBm, and then there is a slight decrease in performance as the UE transmission power increases further. This is because, after 40 dBm transmission power is reached, residual co-user interference overpowers noise and becomes the main performance-limiting issue. Async provides a higher sum rate than the sync mode of operation, according to [57] and [53]. Therefore, SE is much higher in the SIC-async scheme compared to SIC-sync due to the higher achievable sum rate in the async mode than in the sync mode. Due to P-IIC's ability to handle the effective variance of the channel, it performs even better than T-SIC async. The ability for users to access the same shared resource opportunistically by using orthogonal or nonorthogonal methods as they find appropriate boosts SE, which is another benefit of the proposed system for the async mode.

Figure 5 shows the SE performance versus the distributed FBS density for the P-IIC async, IIC async [55], T-SIC async, and T-SIC sync schemes. The picture shows that when FBS density increases in async mode signal processing, especially for the IMMSE types, SE increases quickly, regardless of whether it is in the lower or upper regime. This is because interference and residual error have been removed to a large degree. Thus, the sum rate increases, resulting in better SE. The SE response for both T-SIC sync mode and T-SIC async signal processing is very poor because residual interference increases with the increased received signal power of interfering HMA users; however, they produce a significantly good response in the upper regime as higher FBS deployment surpasses these negative effects. The proposed MMSE-IIC scheme, along with M-QAM modulation and HMA, works better than the other three schemes, regardless of whether it is applied in the lower or upper regime. This is because it lowers residual noise variance outside the MMSE unit and fixes erroneous bit streams iteratively at the MMSE unit by PIC. The proposed approach also applies to HetNets, where the macrocell-allocated spectrum has not been successfully used and may result in spectrum holes that can be effectively filled by the deployment of femtocells. It is important to note that HetNet allows the implantation of femtocells in both licensed and unlicensed spectra. This contributes significantly to raising the SE of the proposed scheme as well. However, in the upper FBS density regime, the IIC async approach [55] performs slightly better than the T-SIC sync and T-SIC async approaches, as it cannot handle residual error properly. However, the use of SISO in place of MIMO does not help the IIC async approach [55] greatly in reaching a higher SE.

The optimum user count and performance of communication networks are significantly impacted by changes to the MA strategy. Figure 6(a) depicts the EE performance in the P-IIC async mode of signal processing versus various BS antennas and UEs. To enhance the EE of the model, it is important to increase the counts of both antennas and UEs. The figure demonstrates that for a certain number of UE devices, the optimal EE can only be reached with a certain number of antennas, and adding more antennas will not make a large difference in performance.

Figure 6(b) shows the trade-off between power consumption and the hardware-impairment design parameters of different transceiver circuits. The hardware-impairment design parameters are on the x-axis, and the percentage of power consumption is on the y-axis. The performance characteristics of the async and sync modes can be obtained from studying their power consumption maps. The elements, namely, leakage power (C_0), electronic circuit power (C_1), electric circuit power (C_2), finite-impulse-response (FIR) filter coefficient (C_3), and beta coefficient (C_4), are the set of hardware-impairing parameters associated with the mapping to improve EE. Leakage power comprises subthreshold and leakage current. Power distribution across the chip is accomplished by an interconnect, which introduces its own transient behavior and is not an ideal conductor. Variations in space and time also cause changes in the activity across the chip, which in turn cause variations in the current that the power supply network is required to supply.



(a) (b)

FIGURE 6. An optimized EE for P-IIC's asynchronous scheme in terms of the numbers of both BS antennas and scheduled UEs (Figure 6a); power consumption vs. hardware impairment parameters (Figure 6b).

The transistors' subthreshold leakage varies with the power supply voltage applied at each location on the chip. The transistor causes leakage current when it does not shut off entirely. While power dissipation in electric circuits is dependent on impedance mismatches, electronic circuit power is dependent on RF chains. The FIR filter coefficient's power loss is caused by multipliers and switching activities. In relation to radiation loss, the beta coefficient depends on various factors, such as wavelength, substrate thickness, and effective dielectric constant. The proposed scheme outperforms other schemes for all hardware-impairing parameters, which is critical for transceiver circuits in ensuring energy-efficient architecture design, as shown in Figure 6(b). Inexpensive and low-power transceiver circuit designs are essential to making future gMIMO-NOMA communications economically feasible. The power consumption of the IIC-Async mode [55] is less than that of the T-SIC async/sync modes for C_0 and C_3, whereas both the T-SIC async scheme and the scheme in [55] perform equally well in terms of power consumption for C_1. For C_2 and C_4, the T-SIC async scheme performs better than the IIC-Async scheme [55] and the T-SIC sync scheme in the context of power consumption.

Power consumption is also related to EE; therefore, less power consumption will raise the possibility of increasing EE. There is no doubt that the proposed scheme is more energy efficient than the rest of the schemes for two reasons: one is the lower power consumption, as seen from Figure 6(b), and the second is the high data rate due to the consideration of HMA, gMIMO, and the transceiver with QAM. Even though the scheme given in [55] consumes less power for most of the hardware-impairment design parameters compared to T-SIC async/sync schemes, it may not be more energy efficient compared to them, as it cannot provide a higher data rate due to the use of SISO. The use of gMIMO provides a much higher SINR than SISO, which in turn significantly improves the data rate. Gigantic MIMO networks, especially on mobile devices, are using increasing numbers of antennas and other circuitry, so energy-efficient architecture is becoming increasingly important.

V. CONCLUSION

The key ideas of this paper include MMSE-IC-based transceivers and suitable asynchronous signal processing for the gMIMO-HMA HetNet architecture. Because it can handle the asynchronous nature of received signals, eliminate residual interference, adaptively employ shared resources orthogonally or nonorthogonally, and take into consideration small cells in HetNet and M-QAM in the transceiver, the suggested architecture performs better than traditional approaches. The results of the performance investigation for OP, SE, EE, and power consumption caused by hardware impairment are presented.

APPENDIX A
PROOF OF PROPOSITION 1

For a noiseless environment, (1) can be written as

$$s = K_{\chi}^{\chi} F \omega, \tag{47}$$

where the estimation of s can be given by

$$\hat{s} = \hat{K}_{\chi}^{\chi} \hat{F} \hat{\omega}. \tag{48}$$

Furthermore, the PIC efficiency can be defined by

$$\eta_{IC} = 1 - E\left[\frac{|s - \zeta \hat{s}|^2}{|s|^2}\right], \tag{49}$$

where ζ is used to denote the estimation coefficient. Now, by substituting (47) and (48) into (49) and assuming that

$\omega = M$ as each channel uses a portion of the source data, consisting of $\log_2 M$ bits, which is mapped to an M-ary symbol and subsequently released from the RF chain to transmit information, we obtain:

$$\begin{aligned}\eta_{IC} &= 1 - E\left[\frac{|K_x^\chi FM - \zeta \hat{K}_x^\chi \hat{F}\hat{\omega}|^2}{|K_x^\chi FM|^2}\right] \\ &= 1 - E\left[\frac{\begin{array}{l}|K_x^\chi FM - \zeta K_x^\chi F\hat{\omega} \\ +\zeta K_x^\chi F\hat{\omega} - \zeta \hat{K}_x^\chi \hat{F}\hat{\omega}|^2\end{array}}{|K_x^\chi FM|^2}\right] \\ &= 1 - E\left[\frac{\begin{array}{l}(K_x^\chi FM - \zeta K_x^\chi F\hat{\omega})^2 \\ +(\zeta K_x^\chi F\hat{\omega} - \zeta \hat{K}_x^\chi \hat{F}\hat{\omega})^2 \\ +2\zeta K_x^\chi F(M - \zeta\hat{\omega})(K_x^\chi F - \hat{K}_x^\chi \hat{F})\hat{\omega}\end{array}}{|K_x^\chi FM|^2}\right].\end{aligned} \tag{50}$$

(50) can be rewritten with certain mathematical operations as

$$\begin{aligned}\eta_{IC} &= 1 - E\left[\frac{|M - \zeta\hat{\omega}|^2}{|M|^2}\right] \\ &\quad - \zeta^2 E\left[|\hat{\omega}|^2\right] E\left[\frac{|K_x^\chi F - \hat{K}_x^\chi \hat{F}|^2}{|K_x^\chi FM|^2}\right] \\ &\quad - E\left[\frac{2\zeta Re\{(M - \zeta\hat{\omega})(K_x^\chi F - \hat{K}_x^\chi \hat{F})\hat{\omega}\}}{|K_x^\chi FM|^2}\right] \\ &\overset{(V)}{=} 1 - E\left[\frac{|M - \zeta\hat{\omega}|^2}{|M|^2}\right] \\ &\quad - \zeta^2 E\left[|\hat{\omega}|^2\right] E\left[\frac{|K_x^\chi F - \hat{K}_x^\chi \hat{F}|^2}{|K_x^\chi FM|^2}\right] \\ &\simeq 1 - \left[\frac{M^2 - 2\zeta E\left[Re\{\hat{\omega}M\}\right] + \zeta^2 E\left[|\hat{\omega}|^2\right]}{|M|^2}\right] \\ &\quad - \zeta^2 E\left[|\hat{\omega}|^2\right]\epsilon,\end{aligned} \tag{51}$$

where (V) indicates that the last term is neglected, assuming that the normalized MSE (NMSE) is uncorrelated to the channel estimation value and ϵ denotes NMSE as

$$\epsilon = E\left[\frac{|K_x^\chi F - \hat{K}_x^\chi \hat{F}|^2}{|K_x^\chi FM|^2}\right]. \tag{52}$$

After differentiating η_{IC} with respect to ζ, we obtain

$$\frac{d\eta_{IC}}{d\zeta} = \frac{2E\left[Re\{\hat{\omega}M\}\right]}{|M|^2} - \frac{2\zeta E\left[|\hat{\omega}|^2\right]}{|M|^2} - 2\zeta E\left[|\hat{\omega}|^2\right]\epsilon.$$

In an ideal case, $\eta_{IC} = 0$, which can be achieved if $\zeta = 0$, but the value of ζ is not equal to zero in a practical case. Therefore, to find the optimal value of ζ, we use

$$\left.\frac{d\eta_{IC}}{d\zeta}\right|_{\zeta=\zeta^*} = 0. \tag{53}$$

Hence, we obtain ζ^* as

$$\zeta^* = \frac{E\left[Re\{\hat{\omega}M\}\right]}{E\left[|\hat{\omega}|^2\right]\left[1 + M^2\epsilon\right]}. \tag{54}$$

APPENDIX B
PROOF OF PROPOSITION 2

If $D_{u,v}$ can be determined by eliminating the u^{th} row and v^{th} column, then in accordance with [58], the determinant of D with dimension $2\varphi \times 2\varphi$ can be written as

$$\begin{aligned}|D|_{2\varphi} &= det(D_{2\varphi\times 2\varphi}) \\ &= \sum_{v=1}^{2\varphi}(-1)^{2\varphi+v} d_{2\varphi,v} det(D_{2\varphi,v}) \\ &= (-1)^{2\varphi+2\varphi} d_{2\varphi,2\varphi} det(D_{2\varphi,2\varphi}) \\ &\quad + (-1)^{2\varphi+(2\varphi-1)} d_{2\varphi,(2\varphi-1)} det(D_{2\varphi,(2\varphi-1)}) \\ &= (-1)^{2\varphi+2\varphi}(1 + a_2^{-1})|D|_{(2\varphi-1)\times(2\varphi-1)} \\ &\quad + (-1)^{2\varphi+(2\varphi-1)}(1 - \delta_0)|D|_{(2\varphi-1)\times(2\varphi-1)} \\ &= (1 + a_2^{-1})|D|_{(2\varphi-1)\times(2\varphi-1)} \\ &\quad - (1 - \delta_0)(-1)^{2\varphi+(2\varphi-2)}(1 - \delta_0)|D|_{(2\varphi-2)\times(2\varphi-2)} \\ &= (1 + a_2^{-1})|D|_{(2\varphi-1)\times(2\varphi-1)} \\ &\quad - (1 - \delta_0)^2|D|_{(2\varphi-2)\times(2\varphi-2)} \\ &= (1 + a_2^{-1})|D|_{(2\varphi-1)} - (1 - \delta_0)^2|D|_{(2\varphi-2)},\end{aligned} \tag{55}$$

where $d_{2\varphi,v}$ stands for the grid component at the $2\varphi^{th}$ row and v^{th} column of $((\beta^* K_x^\chi F \hat{Z}_{MMSE} F^H [K_x^\chi]^H \hat{Z}_{MMSE}^H)^{-1} + Q)$, $a_1 = \beta^* p_1 g_1^2$ and $a_2 = \beta^* p_2 g_2^2$. Likewise the recursive expression, i.e., the $(2\varphi - 1)^{th}$ term of a Fibonacci Sequence of $|D|_{(2\varphi-1)}$, can be given by

$$|D|_{(2\varphi-1)} = (1 + a_1^{-1})|D|_{(2\varphi-2)} - \delta_0^2|D|_{(2\varphi-3)}. \tag{56}$$

By substituting (56) into (55), we obtain

$$\begin{aligned}|D|_{2\varphi} &= (1 + a_2^{-1})\left\{(1 + a_1^{-1})|D|_{(2\varphi-2)} - \delta_0^2|D|_{(2\varphi-3)}\right\} \\ &\quad - (1 - \delta_0)^2|D|_{(2\varphi-2)} \\ &= (1 + a_2^{-1})(1 + a_1^{-1})|D|_{(2\varphi-2)} \\ &\quad - \delta_0^2(1 + a_2^{-1})|D|_{(2\varphi-3)} - (1 - \delta_0)^2|D|_{(2\varphi-2)} \\ &= \left\{(1 + a_2^{-1})(1 + a_1^{-1}) - (1 - \delta_0)^2\right\}|D|_{(2\varphi-2)} \\ &\quad - \delta_0^2(1 + a_2^{-1})|D|_{(2\varphi-3)} \\ &= \left\{a_1^{-1} + a_2^{-1} + a_1{}^{-1}a_2^{-1} + \delta_0(2 - \delta_0)\right\}|D|_{(2\varphi-2)} \\ &\quad - \delta_0^2(1 + a_2^{-1})|D|_{(2\varphi-3)} \\ &= (k_1 + k_2)|D|_{(2\varphi-2)} - k_1k_2|D|_{(2\varphi-3)},\end{aligned} \tag{57}$$

where $k_1 + k_2$ and k_1k_2 are calculated as

$$k_1 + k_2 = a_1^{-1} + a_2^{-1} + a_1^{-1}a_2^{-1} + \delta_0(2 - \delta_0), \tag{58}$$

$$k_1k_2 = \delta_0^2(1 + a_2^{-1}). \tag{59}$$

Furthermore, by substituting (59) into (58), we obtain the following[7]:

$$k_1^2 - a_1^{-1} + a_2{}^{-1} + a_1^{-1}a_2^{-1} + \delta_0(2-\delta_0)k_1 + \delta_0^2(1+a_2^{-1}) = 0. \tag{60}$$

Hence, the roots of the above equation can be found as stated in proposition 2.

APPENDIX C
PROOF OF PROPOSITION 3

If $\varphi = 2$, then (57) becomes

$$|D| = (k_1+k_2)|D|_2 - k_1k_2|D|_1, \tag{61}$$

where $|D|_1$ and $|D|_2$ can be written as

$$|D|_1 = \left(1 + a_1^{-1}\right), \tag{62}$$

$$\begin{aligned} |D|_2 &= \begin{bmatrix} 1+a_1^{-1} & 1-\delta_0 \\ 1-\delta_0 & 1+a_2^{-1} \end{bmatrix} \\ &= (1+a_1^{-1})(1+a_2^{-1}) - (1-\delta_0)^2 \\ &= k_1 + k_2. \end{aligned} \tag{63}$$

By substituting (62) and (63) into (61), we obtain

$$|D| = (k_1+k_2)^2 - k_1k_2(1+a_1^{-1}). \tag{64}$$

From (30), (32), and (64), we get

$$\begin{aligned} R^*_{async} &= \frac{\varphi}{\varphi+\delta_0}\log_2(a_1a_2) + \frac{1}{\varphi+\delta_0}\log_2|D| \\ &= \frac{\varphi}{\varphi+\delta_0}\log_2(a_1a_2) \\ &\quad + \frac{1}{\varphi+\delta_0}\log_2\left\{(k_1+k_2)^2 - k_1k_2(1+a_1^{-1})\right\}. \end{aligned} \tag{65}$$

This proves proposition 3.

REFERENCES

[1] P. Ahokangas, M. Matinmikko-Blue, and S. Yrjölä, "Envisioning a future-proof global 6G from business, regulation, and technology perspectives," *IEEE Commun. Mag.*, vol. 61, no. 2, pp. 72–78, Feb. 2023.

[2] J. Zhang, J. Fan, J. Zhang, D. W. K. Ng, Q. Sun, and B. Ai, "Performance analysis and optimization of NOMA-based cell-free massive MIMO for IoT," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9625–9639, Jun. 2022.

[3] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 176–183, Oct. 2017.

[4] S. M. R. Islam, M. Zeng, O. A. Dobre, and K.-S. Kwak, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 40–47, Apr. 2018.

[5] M. Li, Z. Yuan, Y. Lyu, P. Kyösti, J. Zhang, and W. Fan, "Gigantic MIMO channel characterization: Challenges and enabling solutions," *IEEE Commun. Mag.*, vol. 61, no. 10, pp. 140–146, Oct. 2023.

[6] T. Ahmad, R. Chai, M. Adnan, and Q. Chen, "Low-complexity heuristic algorithm for power allocation and access mode selection in M2M networks," *IEEE Internet Things J.*, vol. 9, no. 2, pp. 1095–1108, Jan. 2022.

[7]The two solutions of a given quadratic equation, $x^2 + bx + c = 0$, can be expressed as follows: $x = \frac{-b \pm \sqrt{b^2-4c}}{2}$, where x indicates an unknown variable, and b and c denote the linear coefficient and constant coefficient, respectively.

[7] M. Zeng, A. Yadav, O. A. Dobre, and H. V. Poor, "Energy-efficient power allocation for MIMO-NOMA with multiple users in a cluster," *IEEE Access*, vol. 6, pp. 5170–5181, 2018.

[8] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, "On the performance gain of NOMA over OMA in uplink communication systems," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 536–568, Jan. 2020.

[9] A. Benjebbour, Y. Kishiyama, and Y. Okumura, "Field trials of improving spectral efficiency by using a smartphone-sized NOMA chipset," *NTT DOCOMO Tech. J.*, vol. 20, no. 1, pp. 4–13, Jul. 2018.

[10] Z. Yang, Z. Ding, P. Fan, and Z. Ma, "Outage performance for dynamic power allocation in hybrid non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 8, pp. 1695–1698, Aug. 2016.

[11] Z. Q. Al-Abbasi and D. K. C. So, "Resource allocation in non-orthogonal and hybrid multiple access system with proportional rate constraint," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6309–6320, Oct. 2017.

[12] F. Tanaka, H. Suganuma, and F. Maehara, "Hybrid multiple access scheme using NOMA and OMA simultaneously considering user request," in *Proc. 24th Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Dec. 2021, pp. 1–5.

[13] K. Wang, Z. Ding, and W. Liang, "A game theory approach for user grouping in hybrid non-orthogonal multiple access systems," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Sep. 2016, pp. 643–647.

[14] Z. Zhao and W. Chen, "An adaptive switching method for sum rate maximization in downlink MISO-NOMA systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[15] X. Wei, H. Al-Obiedollah, K. Cumanan, Z. Ding, and O. A. Dobre, "Energy efficiency maximization for hybrid TDMA-NOMA system with opportunistic time assignment," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8561–8573, Aug. 2022.

[16] Z. Ding, D. Xu, R. Schober, and H. V. Poor, "Hybrid NOMA offloading in multi-user MEC networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 5377–5391, Jul. 2022.

[17] D.-H. Chen and E.-H. Jiang, "Joint power and time allocation in hybrid NoMA/OMA IoT networks for two-way communications," *Entropy*, vol. 24, no. 12, p. 1756, Nov. 2022.

[18] S. K. Zaidi, S. F. Hasan, and X. Gui, "Evaluating the ergodic rate in SWIPT-aided hybrid NOMA," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1870–1873, Sep. 2018.

[19] L. Zhu, Z. Xiao, X.-G. Xia, and D. O. Wu, "Millimeter-wave communications with non-orthogonal multiple access for B5G/6G," *IEEE Access*, vol. 7, pp. 116123–116132, 2019.

[20] L. Zhang, C. Pan, Y. Wang, H. Ren, and K. Wang, "Robust beamforming design for intelligent reflecting surface aided cognitive radio systems with imperfect cascaded CSI," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 1, pp. 186–201, Mar. 2022.

[21] G. Zhou, C. Pan, H. Ren, K. Wang, M. D. Renzo, and A. Nallanathan, "Robust beamforming design for intelligent reflecting surface aided MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658–1662, Oct. 2020.

[22] V. W. S. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key Technologies for 5G Wireless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[23] Z. Chen, Z. Ding, and X. Dai, "Beamforming for combating inter-cluster and intra-cluster interference in hybrid NOMA systems," *IEEE Access*, vol. 4, pp. 4452–4463, 2016.

[24] J. Wang, Z. Lan, C.-W. Pyo, T. Baykas, C.-S. Sum, M. A. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[25] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. Ben Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3693–3708, May 2019.

[26] A. Hussein, C. Rosenberg, and P. Mitran, "Hybrid NOMA in multi-cell networks: From a centralized analysis to practical schemes," *IEEE/ACM Trans. Netw.*, vol. 30, no. 3, pp. 1268–1282, Jun. 2022.

[27] M. Lu, N. Li, X. Tao, and M. Li, "Time minimization in downlink hybrid NOMA wireless powered communication networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–6.

[28] 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.

[29] R. Sirait, W. Hardjawana, and G. Wibisono, "Performance of downlink NOMA for a massive IoT network over a Nakagami-*m* fading channel with optimized power allocation," *IEEE Access*, vol. 11, pp. 67779–67790, 2023.
[30] M. Zeng, N.-P. Nguyen, O. A. Dobre, and H. V. Poor, "Securing downlink massive MIMO-NOMA networks with artificial noise," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 685–699, Jun. 2019.
[31] X. Li, M. Liu, C. Deng, P. T. Mathiopoulos, Z. Ding, and Y. Liu, "Full-duplex cooperative NOMA relaying systems with I/Q imbalance and imperfect SIC," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 17–20, Jan. 2020.
[32] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Nov. 2013, pp. 770–774.
[33] B. Ning, R. Visoz, and A. O. Berthet, "Physical layer abstraction for turbo coded MIMO systems with LMMSE-IC based turbo equalization," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2013, pp. 310–315.
[34] M. Zhang, S. Ahmed, and S. Kim, "Iterative MMSE-based soft MIMO detection with parallel interference cancellation," *IET Commun.*, vol. 11, no. 11, pp. 1775–1781, Aug. 2017.
[35] S. Chan and S. Kim, "Min-sum soft symbol estimation for iterative MMSE detection," *IEEE Access*, vol. 8, pp. 226240–226247, 2020.
[36] S. Chan, M. Zhang, and S. Kim, "A compact soft symbol estimation for iterative MIMO detection," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1790–1794, Oct. 2020.
[37] Z. Xing, J. Zhou, Z. Ge, G. Huang, and M. Hu, "Recovery of high order statistics of PSK signals based on low-rank matrix completion," *IEEE Access*, vol. 11, pp. 12973–12986, 2023.
[38] X. Chen, C. Cheng, X. Lu, and J. Gong, "Iterative interference cancellation based channel estimation for multi-cell massive MIMO systems," in *Proc. 23rd Asia–Pacific Conf. Commun. (APCC)*, Dec. 2017, pp. 1–6.
[39] M. Choi and J. Kim, "Blind signal classification analysis and impact on user pairing and power allocation in nonorthogonal multiple access," *IEEE Access*, vol. 8, pp. 100916–100929, 2020.
[40] G. M. Kraidy and P. S. Rossi, "Iterative MMSE receivers for multiuser MIMO cooperative systems," in *Proc. IEEE 11th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2010, pp. 1–5.
[41] S. Bittner, E. Zimmermann, and G. Fettweis, "Iterative phase noise mitigation in MIMO-OFDM systems with pilot aided channel estimation," in *Proc. IEEE 66th Veh. Technol. Conf.*, Sep. 2007, pp. 1–5.
[42] S. Suyama, H. Suzuki, K. Fukawa, and J. Izumi, "Iterative receiver employing phase noise compensation and channel estimation for millimeter-wave OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1358–1366, Oct. 2009.
[43] Z. Hajji, K. Amis, and A. A. El Bey, "Iterative receivers for large-scale MIMO systems with finite-alphabet simplicity-based detection," *IEEE Access*, vol. 8, pp. 21742–21758, 2020.
[44] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the *K*-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
[45] C.-E. Chen and W.-H. Chung, "An iterative minmax per-stream MSE transceiver design for MIMO interference channel," *IEEE Wireless Commun. Lett.*, vol. 1, no. 3, pp. 229–232, Jun. 2012.
[46] M. Melvasalo, P. Janis, and V. Koivunen, "MMSE equalizer and chip level inter-antenna interference canceler for HSDPA MIMO systems," in *Proc. IEEE 63rd Veh. Technol. Conf.*, May 2006, pp. 1–5.
[47] M. R. McKay, I. B. Collings, and A. M. Tulino, "Achievable sum rate of MIMO MMSE receivers: A general analytic framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 396–410, Jan. 2010.
[48] M. Al-Imari, P. Xiao, and M. A. Imran, "Receiver and resource allocation optimization for uplink NOMA in 5G wireless networks," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2015, pp. 151–155.
[49] W. Yu, H. Jia, and L. Musavian, "Joint adaptive M-QAM modulation and power adaptation for a downlink NOMA network," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 783–796, Feb. 2022.
[50] R. Fa and L. Zhang, "Generalised grouped minimum mean-squared error-based multi-stage interference cancellation scheme for orthogonal frequency division multiple access uplink systems with carrier frequency offsets," *IET Commun.*, vol. 7, no. 7, pp. 685–695, May 2013.
[51] X. Zou, B. He, and H. Jafarkhani, "An analysis of two-user uplink asynchronous non-orthogonal multiple access systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1404–1418, Feb. 2019.
[52] C. Huang, G. Chen, Y. Gong, P. Xu, Z. Han, and J. A. Chambers, "Buffer-aided relay selection for cooperative hybrid NOMA/OMA networks with asynchronous deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2514–2525, Aug. 2021.
[53] M. Ganji, X. Zou, and H. Jafarkhani, "Asynchronous transmission for multiple access channels: Rate-region analysis and system design for uplink NOMA," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4364–4378, Jul. 2021.
[54] S. Kim, J. Kim, and D. Hong, "A new non-orthogonal transceiver for asynchronous grant-free transmission systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1889–1902, Mar. 2021.
[55] M. Sigmund, R. Bomfin, M. Chafii, A. Nimr, and G. Fettweis, "Waveform design for power-domain asynchronous NOMA," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2022, pp. 1–5.
[56] Q. Sun, S. Han, I. Chin-Lin, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, Aug. 2015.
[57] S. Li, Z. Wei, W. Yuan, J. Yuan, B. Bai, D. W. K. Ng, and L. Hanzo, "Faster-than-Nyquist asynchronous NOMA outperforms synchronous NOMA," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1128–1145, Apr. 2022.
[58] D. Poole, *Linear Algebra: A Modern Introduction*, 3rd ed. Boston, MA, USA: Cengage Learning, 2011, p. 803.

JOYDEV GHOSH received the B.Tech. degree in electronics and communication engineering from MAKAUT (formerly WBUT), in 2008, and the M.Tech. degree in telecommunication engineering from the National Institute of Technology (NIT) Durgapur, India, in 2013. He is currently pursuing the Ph.D. degree in computer science and information technology from the Department of Computer Science (CS), College of Computing, Khon Kaen University (KKU), Thailand. His research interests include artificial intelligence (AI), machine learning (ML), deep neural networks (DNNs), the Internet of Things (IoT), orthogonal multiple access (OMA), nonorthogonal multiple access (NOMA), blockchain, small cells, and signal processing.

CESAR VARGAS-ROSALES (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in communications and signal processing and electrical engineering from Louisiana State University. He has also been a Faculty Advisor of the IEEE-HKN Lambda Rho Chapter with Tecnológico de Monterrey. He is the coauthor of the books *Position Location Techniques and Applications* (Academic Press/Elsevier) and *Radio Wave Propagation in Vehicular Environments* (IET). His research interests include personal communications, 5G/6G, cognitive radio, MIMO systems, stochastic modeling, interference, intrusion/anomaly detection, position location, satellite communications, and error-correcting codes. He is a member of the Mexican National Researchers System (SNI), the Mexican Academy of Science (AMC), and the Academy of Engineering of Mexico (AIM). He is a Senior Member of the IEEE Communications Society. He was the Monterrey Chapter Chair and the Technical Program Chair of the IEEE Wireless Communications and Networking Conference (IEEE WCNC). He serves as an Associate Editor for IEEE Access and the *International Journal of Distributed Sensor Networks* and a Guest Editor for the *Applied Sciences* journal.

LUCIANO LEONEL MENDES (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Inatel, Brazil, in 2001 and 2003, respectively, and the Ph.D. degree in electrical engineering from Unicamp, Brazil, in 2007. Since 2001, he has been a Professor with National Telecommunications Institute—Inatel. From 2013 to 2015, he was a Visiting Researcher with the Vodafone Chair Mobile Communications Systems, Technical University of Dresden, where he developed the postdoctoral program sponsored by CNPq. Since 2015, he has been a Research Coordinator with the Radiocommunication Reference Center, Inatel, which is a research project aiming at addressing the main challenges for Brazilian Society regarding mobile, satellite, and terrestrial communications. Since 2016, he has also been a CNPq level two research fellow. In 2017, he was elected as a Research Coordinator of the 5G Brazil Project, an association involving industries, telecom operators, and academia, which aims to fund and build an ecosystem to support 5G, enriching the discussion of the needs in Brazil regarding this networks and how Brazil can contribute to international standardization. He has led several research projects funded by FAPEMIG, FINEP, and BNDES. His main research interest includes solutions for the physical layer of the future fifth generation of mobile communication systems (5G). He has published several papers in the research field in recent years.

IN-HO RA (Member, IEEE) received the Ph.D. degree in computer engineering from Chung-Ang University, Seoul, South Korea, in 1995. From February 2007 to August 2008, he was a Visiting Scholar with the University of South Florida, Tampa, FL, USA. He has been with the School of Computer, Information and Communication Engineering, Kunsan National University, where he is currently a Professor. His research interests include wireless ad hoc and sensor networks, blockchain, the IoT, PS-LTE, and microgrids.

VAN NHAN VO received the B.S. degree in computer science from the University of Da Nang, Da Nang, Vietnam, in 2006, the M.S. degree in computer science from Duy Tan University, Da Nang, in 2014, and the Ph.D. degree in computer science from Khon Kaen University, Thailand, in 2019. He is currently a Lecturer with Duy Tan University. His research interests include information security, physical layer secrecy, radio-frequency energy harvesting, non-orthogonal multiple access, wireless sensor networks, the Internet of Things, unmanned aerial vehicles, and the security of other advanced communication systems.

PHET AIMTONGKHAM received the B.S. degree in computer science and the M.S. and Ph.D. degrees in information technology from the Department of Computer Science, Khon Kaen University, Thailand, in 2013, 2017, and 2020, respectively. He is currently a Lecturer and a Researcher with the Department of Computer Science, College of Computing, Khon Kaen University. His research interests include computer networking, the Internet of Things (IoT), multimedia networks, cybersecurity, and machine learning.

CHAKCHAI SO-IN (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in computer engineering from Kasetsart University, Bangkok, Thailand, in 1999 and 2001, respectively, and the M.S. and Ph.D. degrees in computer engineering from Washington University, St. Louis, MO, USA, in 2006 and 2010, respectively. He is currently a Professor of computer science with the Department of Computer Science, Khon Kaen University, Khon Kaen, Thailand. He has interned with the Cisco Networking Academy (CNAP-NTU, SG), Cisco Systems (Silicon Valley, USA), WiMAX Forums (USA), and Bell Labs (Alcatel-Lucent, USA). He has authored or coauthored over 100 international (technical) publications, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, IEEE/CAA, *IEEE Communications Magazine/IEEE Wireless Communications Magazine*, IEEE INTERNET OF THINGS JOURNAL, IEEE SYSTEMS JOURNAL, *JNCA*, *COMNET*, *MONET*, and *ESWA*; and ten books, including *Mobile and Wireless Nets with IoT*, *Computer Network Lab*, and *Network Security Lab*. His research interests include computer networking and the internet, wireless and mobile networking, the Internet of Things, wireless sensor networks, signal processing, cybersecurity, cyber-physical systems, and applied intelligent systems. He is a Senior Member of ACM. He has served as a Committee Member/reviewer for many journals/publishers, such as IEEE, Elsevier, Springer, Wiley, IET, Inderscience, IEICE, and ETRI, and conferences, such as GLOBECOM, ICC, VTC, WCNC, ICNP, ICNC, and PIMRC. He has served as an Associate Editor for IEEE ACCESS, *PLOS ONE*, *Wireless Networks*, *PeerJ (CS)*, and *ECTI-CIT*.