

Received September 3, 2021, accepted September 20, 2021, date of publication September 29, 2021, date of current version October 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3116613

# Over-the-Air Implementation of NOMA: New Experiments and Future Directions

YUE QI<sup>1</sup>, (Graduate Student Member, IEEE),  
XINLIANG ZHANG<sup>1</sup>, (Graduate Student Member, IEEE), AND  
MOJTABA VAEZI<sup>1</sup>, (Senior Member, IEEE)

Department of Electrical and Computer Engineering, Villanova University, Villanova, PA 19085, USA

Corresponding author: Mojtaba Vaezi (mvaezi@villanova.edu)

This work was supported by Villanova University's Falvey Memorial Library Scholarship Open Access Reserve (SOAR) Fund.

**ABSTRACT** Non-orthogonal multiple access (NOMA) is widely recognized to increase the number of users and enhance the spectral efficiency in fifth-generation (5G) wireless networks and beyond. NOMA is still in the theoretical analysis and simulation phases and fewer experimental works are reported to date. In this paper, we design and implement NOMA in software-defined radio, and evaluate its performance. This includes the real-time realization of the key components of NOMA, i.e., superposition and successive interference cancellation. The main novelty of this paper is to introduce constructing superimposed signals with *varying symbol rates* to enlarge the achievable rate region of the experimented NOMA. By applying varying symbol rates, the set of possible transmission rate pairs enlarges and we can reach higher data rates compared to existing modulation and coding schemes (MCS). We also propose an algorithm to efficiently find the rate pairs. Simulations and experiments demonstrate that NOMA with a varying symbol rate not only can reach higher data rates than orthogonal schemes such as time division multiple access, but it can also outperform existing MCS-based methods which have a fixed symbol rate. The experiments also show that there is a noticeable gap between NOMA in theory and practice. In addition to the new NOMA experiments, we review the state-of-the-art in experimental NOMA. We also discuss several directions for future experiments that can help bridge the gap between theory and practice and bring NOMA to practical communication systems.

**INDEX TERMS** NOMA, superposition coding, successive interference cancellation, software-defined radio (SDR), USRP.

## I. INTRODUCTION

Non-orthogonal multiple access (NOMA) in the downlink, also known as power-domain NOMA, is a promising solution for increasing the number of users and enhancing the spectral efficiency of communication systems [1]. Unlike orthogonal multiple access (OMA) techniques that serve only one user per time/frequency resource block, in theory, NOMA can serve an arbitrary number of users in each resource block. For this reason, NOMA has appeared as an appealing solution for increasing connectivity and improving the spectral efficiency in fifth-generation (5G) cellular networks and beyond [1]–[6].

NOMA exploits the difference of channel gains between multiplexed users. The key techniques for single-input and

single-output (SISO)-NOMA are *superposition coding* (SC) at the transmitter and *successive interference cancellation* (SIC) at the receiver. SIC is used at receivers with a stronger channel to decode and cancel interference from all weaker users before decoding its desired information. If  $K$  users are multiplexed to use one resource block, and the users are sorted based on their channel strength,  $i$ th user should apply interference cancellation  $i - 1$  times to decode and remove the signal of all users who have a weaker channel gain. Each interference cancellation will incur some error and this error propagation will increase the bit error rate (BER). In this sense, two-user SISO-NOMA is the simplest case in which only one interference cancellation is required.

While NOMA has attracted significant attention for academic research, an important question to move forward is whether this method will enjoy the theoretically-promised gains over OMA in practice. Generally, although NOMA has

The associate editor coordinating the review of this manuscript and approving it for publication was Miguel López-Benítez<sup>1</sup>.

been under investigation in academia, standardization bodies [7], and industry [8], the research is still mostly in the theoretical analysis and simulation phases [5], [6], [8]. Attempts on the experimental validation and feasibility analysis of NOMA in real wireless environments remain limited. Some pioneering work can be found in [2], [9]–[11]. There are still several open questions and challenges on practical implementation issues, which are the main interest of this article. Software-defined radio (SDR) [12] is considered a valuable tool to evaluate the performance of practical systems. One can flexibly reconfigure an SDR and implement communication modules in software instead of hardware. Universal software radio peripheral (USRP) developed by National Instruments (NI) [13] is a type of SDR that is commonly used by research labs and universities.

This article is meant to serve as a reference for experimental NOMA. We have used the NI USRP-2974 [13] with LabVIEW communication system design suite [14], [15] to set new experiments for the two-user SISO-NOMA to enhance spectral efficiency and shed more light on when and to what extent NOMA can improve spectral efficiency in practice. Besides, we have summarized existing literature on experimental NOMA and discussed challenges and future directions.

We have both simulations and experiments for over-the-air transmission of the two-user NOMA system. The contributions of this paper can be summarized as follows:

1. One novelty in the designed simulation and experiment is the introduction of varying symbol rate transmission to improve the achievable rates. By varying the symbol rate of each NOMA user, the solution space is enlarged compared to only applying the existing modulation and coding scheme (MCS). This, in turn, helps to achieve a higher rate region for NOMA users. In experiments, practical imperfectness such as synchronization, channel estimation, SIC, NOMA header transmission is included.
2. It is well-known that NOMA reaches higher rates than OMA in theory. However, in this paper, we show this is not always the case in practice. Specifically, the experiment shows that NOMA outperforms OMA only in asymmetric channels. However, in symmetric channels where the channel gains of the two users are very close, the NOMA rate region could even be worse than that of OMA for various reasons, including SIC error and imperfect channel state information.
3. We propose an algorithm to more efficiently find the rate pairs for implementation. The algorithm first finds OMA rate pairs for each user based on the code tables with varying symbol rates and then uses a bisection search to find the feasible NOMA rate pairs for both users. The algorithm is more time-saving compared to Monte Carlo simulations over all possible rate pairs.
4. Unlike several existing results, e.g., [9], [16], where the modulation, coding rate, and/or power allocation information are assumed to be known by the receivers

and sometimes a cable connection is used for stable power control, in our experiments, we transmit this information, wirelessly which is more practical. Moreover, we provide various details of the design which is not always reported in the literature, hoping to be useful for the reproduction of the results. This includes preamble and NOMA header which contains power allocation factor, adaptive MCS, symbol rate, etc. NOMA header is designed in detail and transmitted through the air.

The remainder of the paper is organized as follows. We divide the paper into three main parts.

1. Existing literature: We first review existing experimental results in Section II and discuss practical challenges in Section III. In Section IV, we describe the theory of NOMA.
2. New experiments development: The implementation of NOMA is introduced in Section V, in which we discuss the details of our designed experiment and transmission protocols. We present the concept of varying symbol rates by simulation in Section VI and experimental results in Section VII.
3. Future directions and outlook: We summarize future directions in Section VIII, and we conclude the paper in Section IX.

## II. PRIOR EXPERIMENTAL VALIDATIONS OF NOMA

The earliest work on the implementation of downlink two-user SISO-NOMA can be traced back to [9], where the authors show SC can provide significant spectral efficiency gain over *time-division multiplexing*. The experiment is designed using *orthogonal frequency division multiplexing* (OFDM) and the IEEE 802.11a parameters, and achievable rates are measured using the MCS [17]. The USRP boards of the transmitter (Tx) and user equipments (UEs) are connected via cables.

NOMA combined with multiple-input multiple-output (MIMO) has been investigated by Benjebbour *et al.* from NTT DOCOMO in [2], [18]–[20]. The benefits of two-user NOMA over OMA with a two-transmit and two-receive antennas system ( $2 \times 2$ ) are evaluated in a system-level simulation and experimental trials in [2]. In [19], [20], a setup of an outdoor experimental trial using NOMA with open-loop  $2 \times 2$  MIMO is introduced and closed-loop  $4 \times 2$  MIMO with multiple users is investigated in [18], [19]. The results show that a significant difference channel gain can increase the performance gain, and the user pairing and feedback in the closed-loop case are important for NOMA performance.

Xiong *et al.* [11] and Wei *et al.* [10], [21] implement a few SISO-NOMA testbeds. The offline and real-time experiments are considered. In the offline experiment, USRP is treated as the radio frequency (RF) terminal. The decoding and SIC are processed offline. In a real-time experiment, the time complexity of real-time SIC has been evaluated in which the SIC procedure requires around twice the time

**TABLE 1.** Summary of existing NOMA implementation.

Reference	Year	Setting	Equipment Type	Technical Contribution
[9]	2012	2-UE SISO	USRP	Implementing SC and SIC in SISO system over Gaussian broadcast channel (BC).
[2]	2015	2-UE MIMO	DOCOMO MediaTek Helio	System-level simulation and experimental trials.
[19]	2018	2, 3, 4-UE MIMO		Implementing 4 × 2 MIMO-NOMA with multiple users in outdoor environments.
[20]	2019	2-UE MIMO		Implementing two-user 2 × 2 MIMO-NOMA in outdoor environments.
[10]	2015	2-UE SISO	USRP B200	Offline implementation to evaluating the time cost for the real-time experiments.
[11]	2016		USRP B200	Implementing upper layers and evaluating the throughput for NOMA and OMA.
[21]	2017		USRP B210	Upgrading bandwidth from 5MHz to 10MHz, and evaluating throughput loss.
[22]	2017	3-UE MISO	USRP N210	Exploiting network coding in NOMA to increase throughput.
[24]	2017	2-UE SISO	USRP B210 USRP 2920	Offline experiments for uplink NOMA. Designing uplink preamble for each user.
[23]	2019	2-UE SISO	USRP 2920	Experimenting NOMA over downlink shared channel based on 3GPP standard.
[25]	2020	2-UE SISO 2, 3-UE MISO	USRP N210	Designing precoders for multi-antenna systems in PHY and MAC.
[26]	2020	2, 3-UE SISO	USRP	Analyzing and validating the BER of NOMA using QPSK modulation.
[27]	2020	4-UE SISO	USRP 2943	Implementing a four-user NOMA system.
[28]	2020	3, 4-UE SISO		Analyzing and validating the BER of NOMA using BPSK.
[29]	2018	2-UE SISO	USRP 2944	Introducing NOMA to WiFi technology.
[30]	2020			Designing downlink NOMA frame structure in IEEE 802.11 standards.
[31]	2020			Using constellation rotation to migrate phase noise.
[32]	2020	4-UE SISO	USRP N210	Evaluating the BER of NOMA system after applying SIC.
[33]	2020	2-UE SISO	USRP 2922	The channel estimation and quantization impacts on NOMA-OFDM.
[34]	2021	2-UE SISO	USRP	Implementing cooperative NOMA with one relay node.

as direct decoding [10]. Paper [11] includes upper-layer protocols in the SISO-NOMA system and the throughput of NOMA and OMA has been compared there. Later, the authors increased their testbed bandwidth from 5MHz to 10MHz and evaluated the throughput loss [21]. The authors in [22] have exploited network coding in NOMA. The scheme is named network-coded multiple access, which jointly uses network coding and multi-user decoding to boost NOMA throughput. Paper [23] has realized NOMA over downlink shared channels according to the Third Generation Partnership Project (3GPP) standard. Specifically, low-density parity-check (LDPC) channel coding instead of Turbo coding is applied. The simulation and an offline experiment show the feasibility of NOMA but also indicates a high complexity implementation in SDR. An uplink NOMA implementation is presented in [24], where the preamble frame structure for each uplink user is designed.

Very recently, a multiple-input and single-output (MISO) NOMA system ( $2 \times 1$ ) with two and three users is considered in [25]. Linear precoders are designed in PHY and medium access control (MAC) protocols, and a SIC scheme is proposed using reference signals for interference subtraction and signal detection, which can achieve higher weighted sum rates in NOMA networks compared with zero-forcing SIC. Paper [26] studies the performance of the quadrature phase-shift keying (QPSK) modulation over the downlink NOMA system and derives and validates the error probability on a USRP SDR platform. A four-user NOMA network is implemented in [27], [28], [32]. The bit error behavior of binary phase-shift keying (BPSK) for NOMA has been analyzed and validated in the SDR platform [28]. The BER loss is evaluated with and without SIC [32]. Khorov *et al.* introduced the IEEE 802.11 WiFi standards into NOMA [29]–[31], [35]. Specifically, in [29], they have validated the WiFi network in a

slave computer (e.g., field-programmable gate arrays (FPGA) board) rather than a master computer (CPU in the host), which is more efficient than offline experiments. In [30], modifications of the physical (PHY) layer frames and the medium access control (MAC) protocols for downlink transmissions are considered based on IEEE 802.11 standards. To decrease the influence of phase noise in WiFi NOMA systems, [31], [35] extend their NOMA WiFi prototype to enable constellation rotation. The channel estimation and channel quantization impacts on NOMA OFDM have been investigated in [33] with varying pilot-to-data ratios. It reflects that with long channel estimation pilots the BER performance can be better. A dedicated relay-cooperative NOMA implementation is explored in [34] to enhance the performance of the weak user.

Overall, the study in [9]–[11], [25] concludes that NOMA can improve system performance compared with OMA. Besides, papers [9], [10], [23], [32] also indicate that real-time implementation and SIC requires a high complexity. The summary of technical contributions to NOMA implementation is shown in Table 1. One main difference between the existing literature and our work is that we proposed and use a varying symbol rate in this paper. Besides, we design a detailed NOMA header and transmit it over the air. In [9], [16], the modulation, coding rate, and power allocation information are assumed to be known by the receiver, and a cable is connected for stable power control. In contrast, in this paper, we transmit the users' information wirelessly which is more practical.

### III. PRACTICAL CHALLENGES

There are several implementation challenges in experimental NOMA. We mainly elaborate on synchronization and SIC in the following.

## A. SYNCHRONIZATION

Synchronization is crucial in achieving reliable and robust communication system performance. In long-term evolution (LTE) [37], there are two downlink synchronization reference signals, a primary synchronization signal (PSS) for symbol timing and cell search, and a secondary synchronization signal (SSS) for frame timing. In IEEE 802.11 wireless local access network (WLAN) standards [17], all synchronization is realized within symbol sequences of legacy preambles, i.e., short training field (L-STF) for packet detection and carrier frequency offset (CFO) correction, and long training field (L-LTF) for fine synchronization and channel estimation [38]. There are three major synchronization functions in PHY [39], as described below.

### a: FRAME SYNCHRONIZATION

*Frame synchronization*, also known as *initial acquisition*, is the first and coarse task through which a receiver (Rx) establishes a communication link with the Tx and calibrates its parameters. Employing an external clock module [26]–[28] can help coarsely synchronize multiple devices.

### b: TIME SYNCHRONIZATION

This is needed to achieve a fine time and/or frequency synchronization between the Rx sampling clock and the Tx channel symbol clock. In [26]–[28], symbol timing synchronization is desired for a single-carrier system. In NOMA-OFDM systems [25], [34], time and frequency synchronization can be realized by the Schmidl-Cox algorithm [36]. Briefly, it requires inserting a known preamble at the beginning of transmission, and receivers detect them periodically via the correlation and accumulation.

### c: CARRIER SYNCHRONIZATION

This concerns the estimation and compensation of the carrier frequency and phase differences between the Tx and Rx oscillators. The CFO synchronization is necessary for uplink NOMA [24] since CFO leads to multiple access interference. The users use the auto-correlation of the PSS in LTE to perform synchronization [24], [40]. Table 2 shows the synchronization types and technologies employed in NOMA papers.

## B. SUCCESSIVE INTERFERENCE CANCELLATION

### d: IMPERFECT SYNCHRONIZATION AND CHANNEL ESTIMATION

Timing mismatch or imperfect channel estimation will result in *residual* cancellation errors [41], and finally imperfect SIC. The synchronization errors may come from pulse-shaping and the alignment of symbols. On the other hand, if the strong user's channel state information (CSI) is not estimated perfectly, then the weak user's signal cannot be reconstructed correctly from the superimposed signal.

### e: SIC ERROR PROPAGATION

Once decoding error occurs to a user, it is quite likely that all subsequent users in the SIC decoding order will be decoded

TABLE 2. Synchronization in NOMA implementation.

Reference	Types	Technologies and Methods
[9] [16]	a, b	The preamble sequence is used for time and frequency synchronization by repeating a pseudo-random training sequence.
[24]	b, c	The PSS of LTE is first adopted for CFO. Time synchronization preamble is sent from UEs to Tx for uplink transmission timing.
[25] [30]	a, b, c	Synchronization is realized by IEEE 802.11 legacy preambles, i.e., L-STF and L-LTF.
[26]	a, b	USRPs have synchronized clocks due to an external OctoClock. The Tx transmits a set of given symbols for synchronization and channel equalization.
[27] [28]	a, b, c	A clock distribution module (CDA 2990) is used for coarse synchronization. Pilot symbols are used for sample timing, CFO correction, and channel estimation.
[34]	b	The preamble includes known sequences. The method is the Schmidl-Cox algorithm [36].
This work	a, b	Trigger operation for frame acquisition. Time synchronization by threshold detection.

improperly as well. There are several ways to lower the error propagation. If the number of users is small enough, this effect can be mitigated by utilizing stronger codes or longer blocks [41]. The choice of different SIC schemes can also affect the performance of BER and achievable rate in practice, for example, codeword-level SIC can reduce BER compared with symbol-level SIC [42]. The main difference between the two methods is that channel decoding and re-encoding are included in codeword-level SIC while not in symbol-level SIC.

## IV. NOMA: FROM THEORY TO PRACTICE

We summarize the system model achieving the capacity of the SISO-NOMA using SC and SIC. To realize the superimposed signal, signal reconstruction, and interference cancellation in NOMA receivers, we build the NOMA modules in LabVIEW communication suites system based on Fig. 1.

### A. SUPERPOSITION CODING

As illustrated in Fig. 1, the Tx generates two independent random data streams for UE1 and UE2 following a Bernoulli distribution with  $p = \frac{1}{2}$ . The data of two paired UEs are separately encoded, modulated into complex-valued symbols using modulation types such as BPSK, QPSK, 16QAM, and filtered. In general, UE1 upsamples with  $k_1$  samples per symbol and applies pulse shaping filter  $p_1(m)$ , whereas UE2 upsamples with  $k_2$  samples per symbol and applies pulse shaping filter  $p_2(m)$ , in which  $m$  denotes the input index. Denote the signal sequence for each user per channel use as  $s_i$ ,  $i = \{1, 2\}$ . Then, a fraction  $\alpha \in [0, 1]$  of the power  $P$  is assigned to UE1, and the rest of the power is allocated to UE2. The signal of two users with different power are superposed in the power domain, which can be constructed as

$$s = \sqrt{\alpha} s_1 + \sqrt{1 - \alpha} s_2 \quad (1)$$

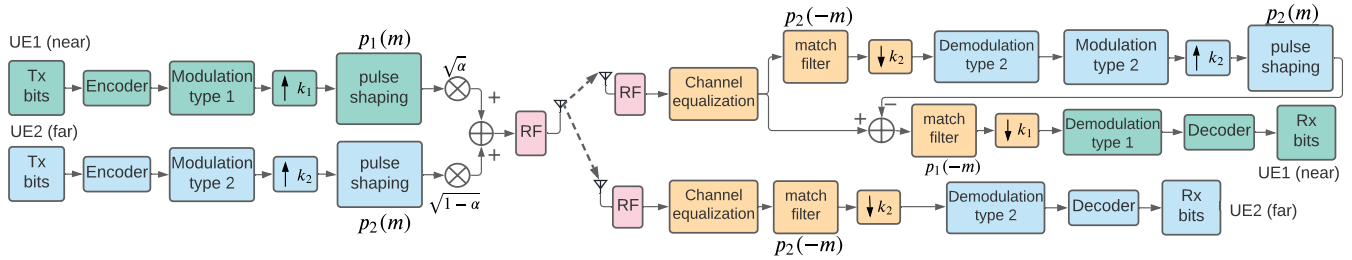


FIGURE 1. Two-user SISO-NOMA system configuration used in this paper.

The superimposed signal is then sent out through USRP with radio frequency (RF) 3GHz. To elaborate on how SC is performed, schematic diagrams are given in Fig. 2. When  $\alpha$  is small (Fig. 2(a) to Fig. 2(c)), the QPSK constellation of UE1 with smaller power is superposed on that of UE2 with a higher power. The super-constellation of the superimposed signal performs as a 16 quadrature amplitude modulation (QAM). When  $\alpha$  is large (Fig. 2(d) to Fig. 2(f)), UE1 chooses QPSK, and UE2 chooses BPSK, the super-constellation performs as an 8PSK constellation with possible constellation points overlapping.

**B. SUCCESSIVE INTERFERENCE CANCELLATION**

Denote the complex channel gains for UE1 and UE2 as  $h_1$  and  $h_2$ , and assume, without loss of generality,  $|h_1|^2 > |h_2|^2$ . The received signal at each user is:

$$y_1 = h_1s + n_1 = h_1\sqrt{\alpha}s_1 + h_1\sqrt{1-\alpha}s_2 + n_1, \quad (2a)$$

$$y_2 = h_2s + n_2 = h_2\sqrt{\alpha}s_1 + h_2\sqrt{1-\alpha}s_2 + n_2, \quad (2b)$$

where  $n_i \sim \mathcal{CN}(0, N_0)$ ,  $i = 1, 2$  denotes the additive white Gaussian noise (AWGN) at each user. UE1 has a stronger channel to Tx, and hence can always decode the information which can be decoded at UE2. The key steps are [4]:

- The weaker UE (UE2) estimates its channel and decodes its data from  $y_2$  by treating  $s_1$  as Gaussian noise. Let  $\hat{h}_2$  be UE2's estimated channel. The recovered signal of UE2 is

$$\hat{s}_2 = (\hat{h}_2\sqrt{1-\alpha})^{-1}y_2. \quad (3)$$

The signal then is filtered using the matched filter  $p_2(-m)$ , downsampled with  $k_2$ , and demodulated using modulation type 2 to recover UE2's data streams.

- At the stronger UE (UE1), four steps are required to recover its message from the received signal  $y_1$ :
  - Estimate CSI ( $\hat{h}_1$ ).
  - Filter, decode, and demodulate the signal of UE2 from  $y_1$ , then reconstruct UE2's signal using modulation type 2 and  $k_2$  to obtain  $\hat{s}_2$ .
  - Subtract  $\hat{h}_1\sqrt{1-\alpha}\hat{s}_2$  from the received signal  $y_1$ :

$$\begin{aligned} \hat{y}_1 &= y_1 - \hat{h}_1\sqrt{1-\alpha}\hat{s}_2 \\ &= h_1\sqrt{\alpha}s_1 + \sqrt{1-\alpha}(h_1s_2 - \hat{h}_1\hat{s}_2) + n_2. \quad (4) \\ &= 0, \text{ if perfect CSI and } \hat{s}_2=s_2 \end{aligned}$$

- Decode and demodulate the signal of UE1 from  $\hat{y}_1$  and obtain the estimated signal  $\hat{s}_1$ .

The Tx needs to share the knowledge of the power allocation efficient  $\alpha$  and modulation types with both users so that they can realize SIC and decode their signals. In the example of Fig. 2(a) to Fig. 2(c), UE1—with a stronger channel—can decode the QPSK for UE2 first, and then subtracts it from the superimposed signal to obtain its QPSK. UE2—with a weaker channel—treats UE1's signal as noise and decodes its message by considering the superimposed 16QAM as a QPSK. In the example of Fig. 2(f), the 8QAM super-constellation is cross overlapped. To successfully decode and satisfy BER requirements, one approach is to avoid constellation cross overlapping and erroneous decision and assign suitable power allocation [43], [44].

In practice, although channel coding can be used to ensure that signals are decoded correctly, imperfect synchronization may cause an imperfect SIC at UE1 and will degrade the performance of NOMA. A more detailed analysis of imperfectness has been discussed in Section III.

**C. ACHIEVING RATE PAIRS**

The achievable data rates for UE1 and UE2 in NOMA are given by [41]

$$r_1 \leq \log\left(1 + \frac{\alpha P|h_1|^2}{N_0}\right), \quad (5a)$$

$$r_2 \leq \log\left(1 + \frac{(1-\alpha)P|h_2|^2}{\alpha P|h_2|^2 + N_0}\right), \quad (5b)$$

in which  $P$  represent the total power. The signal-to-interference-plus-noise ratio (SINR) of UE2 in (5b) is given by:

$$\text{SINR} = \frac{(1-\alpha)P|h_2|^2}{\alpha P|h_2|^2 + N_0} \quad (6)$$

The *capacity region* of the NOMA channel is obtained by varying  $\alpha$  from 0 to 1 to get all possible transmission rate pairs  $(r_1, r_2)$ . It is worth noting that when all power is allocated to UE1 ( $\alpha = 1$ ), the channel reduces to the point-to-point (P2P) transmission between the Tx and UE1, which implies an OMA scheme since only UE1 is served and  $r_2 = 0$ . When  $\alpha = 0$ , only UE2 is active in the network and  $r_1 = 0$ . In the OMA case, assume time division multiple access (TDMA),

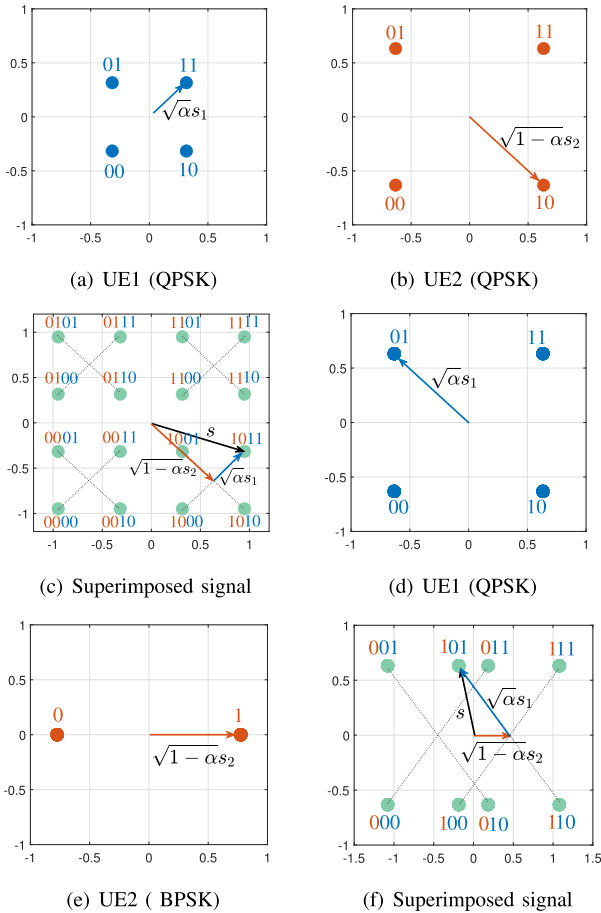


FIGURE 2. Depiction of the super-constellations for  $\alpha = 0.2$  in (a)-(c) and for  $\alpha = 0.8$  in (d)-(f).

the achievable rate region for UE1 and UE2 is written by [45], [46]

$$r_1 \leq \tau \log\left(1 + \frac{P|h_1|^2}{N_0}\right), \quad (7a)$$

$$r_2 \leq (1 - \tau) \log\left(1 + \frac{P|h_2|^2}{N_0}\right), \quad (7b)$$

in which a fraction  $\tau \in [0, 1]$  of time is dedicated to UE1, and the remaining fraction  $1 - \tau$  of the time is dedicated to UE2. With NOMA, the entire time or bandwidth is simultaneously shared by two users. However, with OMA, UE1 uses part of time or bandwidth and the remaining resource is assigned to UE2.

## V. NOMA WITH VARYING SYMBOL RATES

### A. PRACTICAL NOMA USING AMC

Adaptive modulation and coding (AMC) is used to adjust the modulation types and coding rates based on the channel conditions. The algorithm to select the optimum MCS is a key for AMC [47]. When the channel conditions are poor, a smaller modulation method and coding rate are used; when the channel conditions are good, a larger modulation method

TABLE 3. MCS parameters in the IEEE 802.11ac [17].

SNR (dB)	[2, 5)	[5, 9)	[9, 11)	[11, 15)	[15, 18)	[18, 20)	[20, 25)
Modulation	BPSK	QPSK	QPSK	16QAM	16QAM	64QAM	64QAM
Coding rate $\rho$	1/2	1/2	3/4	1/2	3/4	2/3	3/4
$\eta = \rho \log_2 M$	1/2	1	3/2	2	3	4	4.5

is favored to maximize the transmission rate. Table 3 shows part of the MCS parameters used in IEEE 802.11ac [17], [48].  $M \in \{2, 4, 16, \dots\}$  denotes the order of states corresponding to the modulation types, BPSK, QPSK, and 16QAM, etc.  $\rho$  is the coding rate, and the coding efficiency  $\eta = \rho \log_2 M$  represents the useful data bits each symbol can carry.

For the OMA case where only one user is being communicated to, the optimal MCS can be selected by comparing the instantaneous signal-to-noise ratio (SNR) to the range of SNR values [47]. In NOMA, MCS adaption for users should be done simultaneously within targeted BERs. Besides, employing MCS adaption based on Table 3 should also consider power allocation and user scheduling [3]. Besides, employing MCS adaption based on Table 3 should also consider power allocation and user scheduling [3]. There are some solutions for employing MCS selection. In [10], a fixed MCS combination (UE1: QPSK, 16QAM, and 64QAM; UE2: QPSK) is selected and the targeted block error rate is satisfied by varying the SNR and power allocation. In [11], [21], suitable MCS schemes of LTE (UE1: 64QAM; UE2: QPSK) are roughly chosen and applied manually according to the average throughput for each UE. In [27], the 4QAM constellation is applied for all users with different channel gains. In [9], [16], MCS adaption is employed based on the SNR for initial reference, and then adjust the power allocation to support targeted packet error rates.

We employ MCS adaption for OMA cases while maintaining the BER below a threshold. In NOMA scenarios, we use the selected MCS in OMA first. Then, we change the MCS to assure the targeted BERs for both users. Table 3 shows part of the MCS parameters used in IEEE 802.11ac. Convolutional coding can be applied to reduce the bit error probability for a noisy channel [49], [50]. However, the system will suffer a remarkable power penalty, coding gain, and rate reduction [51]. Our main attention is to reach as high achievable rate as possible for the NOMA system. Thus, to increase the transmission rate within acceptable BER, we choose to switch between enabling and disabling the convolutional coding.

The MCS provides the code tables to determine a set of largest simultaneously achievable rates at each user. Thus, we can traverse the library to find all possible rate pairs and keep the BER within a practical reliability regime ( $\epsilon \lesssim 10^{-3}$ ) [44], [52], [53].

### B. SYMBOL RATE AND RATE TABLE

In this section, we design three finite libraries of finite-blocklength codes based on finite constellations. The main novelty in building these code libraries, also called code tables, is the introduction of varying symbol rates. These

libraries are named  $C_1$ ,  $C_2$ , and  $C_3$  which include various possible combinations of the transmission parameters. In the following, we describe their construction and differences in detail.

Each code table is built using three independent parameters  $M$ ,  $\rho$ , and  $k$ , where  $M$  denotes the modulation order,  $\rho$  is the channel coding rate (convolutional coding), and  $k$  is the samples per symbol, or oversampling factor—see Table 4 for an example of  $C_1$ . The last column of each code table is the spectral efficiency  $\tilde{r}$  [bps/Hz] evaluated by

$$\tilde{r} = \frac{\rho \log_2 M \cdot R_s}{B}, \tag{8}$$

in which  $B$  is the assigned bandwidth [Hz] and variable  $R_s$  is the symbol rate [symbols/s]. The M-QAM constellations have Nyquist data pulses  $B = R_{s,\max} \geq R_s$  [54]. In digital communication, the sampling frequency  $f_s$  is provided at UEs. We have  $R_s = \frac{f_s}{k}$  and  $R_{s,\max}$  results in the minimum samples per symbol  $k_{\min}$ . Thus, the spectral efficiency can be rewritten as

$$\tilde{r} = \frac{\rho \log_2 M \cdot k_{\min}}{k}. \tag{9}$$

TABLE 4. An example of code table  $C_1$ .

$M$	$\rho$	$k$	$\tilde{r}$
2	1	10	1
2	1	12	5/6
2	1	⋮	⋮
2	1	40	1/4
4	1	10	2
4	1	12	5/3
4	1	⋮	⋮
4	1	40	1/2

We should indicate that we will use  $\hat{r}$  for measured spectral efficiency. Further, all parameters may be used with subscript  $i$ ,  $i \in \{1, 2\}$ , referring to the UE index. For example,  $\tilde{r}_1 = \frac{\rho_1 \log_2 M_1 \cdot k_{\min}}{k_1}$  for UE1.

The difference between code tables  $C_1$ ,  $C_2$ , and  $C_3$  is that in  $C_1$  and  $C_2$  one parameter is fixed whereas in  $C_3$  all parameters are varied. Specifically, in  $C_1$   $\rho$  is fixed to 1—no channel coding—and in  $C_2$   $k$  is fixed to 10.

An example of code table  $C_1$  is shown in Table 4. In this table  $M = \{2, 4\}$  represents modulation mappers BPSK and QPSK,  $\rho = 1$  (no convolutional coding), and samples per symbol  $k$  varies from 10 to 40 with step 2. Thus,  $k = \{10, 12, \dots, 40\}$  and  $k_{\min} = 10$ . The cardinality of  $C_1$  (i.e., possible combinations) is  $|C_1| = |M| |\rho| |k| = 2 \times 1 \times 16$  and thus  $C_1$  has 32 different combinations in this example. The combination of NOMA modulation types in UE1 and UE2 are both BPSK  $\{M_1 = 2, M_2 = 2\}$ ; UE1 and UE2 both QPSK  $\{M_1 = 4, M_2 = 4\}$ ; UE1 is BPSK and UE2 is QPSK  $\{M_1 = 2, M_2 = 4\}$ , UE1 QPSK and UE2 BPSK  $\{M_1 = 4, M_2 = 2\}$ . The solution space of the discrete rate

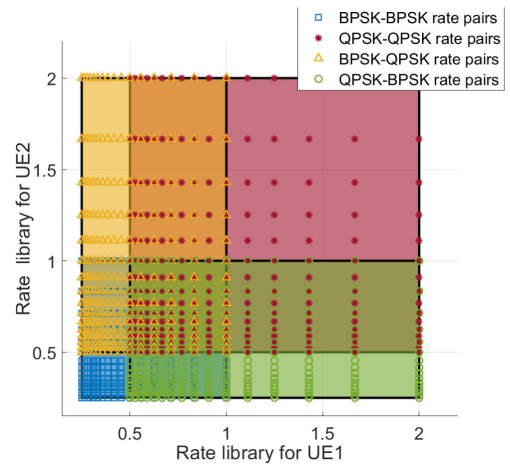


FIGURE 3. Achievable rate pairs using different symbol rates, coding rate, and modulation combinations in code table  $C_1$ .

pairs, i.e.,  $\hat{r}_1$  versus  $\hat{r}_2$ , has the size of  $32 \times 32$  combinations in total, shown in Fig. 3.

The code table  $C_2$ , in which the samples per symbol is fixed whereas the coding rate varies, is built for the sake of comparison with [9], [25]. Specifically, two modulation mappers  $M = \{2, 4\}$  and four coding rates  $\rho = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\}$  are considered. Then all 8 possible pairings of MCS are  $\{\log_2 2, \log_2 4\} \times \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\} = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}, 1, \frac{4}{3}, \frac{3}{2}, \frac{5}{3}\}$ . The code table example can be found in Table 5. The cardinality of  $C_2$  is 8, i.e., there are eight possible achievable rate pairs for one user. Then, the solution space of the rate pairs is  $8 \times 8$  combinations in total, as is shown in Fig. 4 for  $k_1 = k_2 = 10$ . Hence, a fixed sample per symbol may reduce the solution space. Thus, in one frequency carrier, varying the symbol rate can enlarge the solution space.

TABLE 5. An example of code table  $C_2$ .

$M$	$\rho$	$k$	$\tilde{r}$
2	1/2	10	1/2
2	2/3	10	2/3
2	3/4	10	3/4
2	5/6	10	5/6
4	1/2	10	1
4	2/3	10	4/3
4	3/4	10	3/2
4	5/6	10	5/3

Lastly, by combining the varying symbol rate method and MCS with the fixed symbol rate method we create a larger rate table  $C_3$  in which the search is over  $\rho$ ,  $M$ , and  $k$  for both users. The cardinality of this code table is  $|C_3| = |M| |\rho| |k| = 2 \times 5 \times 16 = 160$ . Note that in this case  $\rho = \{1, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\}$ . The corresponding code table example can be found in Table 6. The solution space of the rate pairs in this case is  $160 \times 160$ .

To summarize, three code tables are considered:

- $C_1$  (proposed): Varying symbol rate with fixed coding rate. We use this code table to verify the efficiency of the proposed varying symbol rate method.

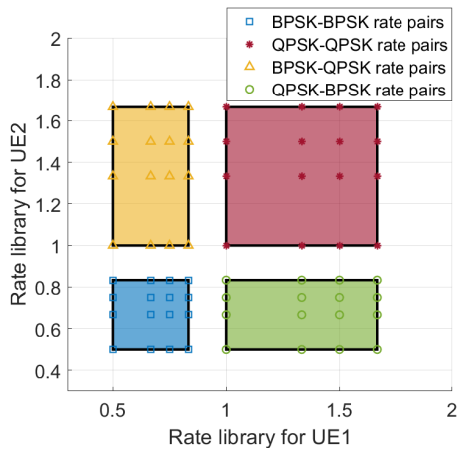


FIGURE 4. Achievable rate pairs only using different coding rate and modulation combinations in code table  $C_2$  [9].

TABLE 6. An example of code table  $C_3$ .

$M$	$\rho$	$k$	$\tilde{r}$
2	1	10	1
2	1/2	10	1/2
2	2/3	10	2/3
2	3/4	10	3/4
2	5/6	10	5/6
2	1	12	5/6
2	1/2	12	5/12
2	2/3	12	5/9
2	3/4	12	5/8
2	5/6	12	25/36
⋮	⋮	⋮	⋮
4	1	40	1/2
4	1/2	40	1/4
4	2/3	40	1/3
4	3/4	40	3/8
4	5/6	40	5/12

- $C_2$ : Fixed symbol rate with varying coding rate. This code table is the existing MCS-based method commonly used in many works [9], [11].
- $C_3$  (proposed): Varying symbol rate and varying coding rate; We combine the two methods to enlarge the solution space.

Given an  $\alpha$  and any code table with finite feasible rate pairs, the problem of finding the set of largest simultaneously achievable rates becomes a search over the feasible rate space. Thus, the problem can be written as following optimization [9]:

$$\max_{\hat{r}_1 \in \tilde{r}} \hat{r}_1 \tag{10a}$$

$$s.t. \hat{r}_2 \in \tilde{r}, \tag{10b}$$

$$BER_1 \leq \epsilon, BER_2 \leq \epsilon, \tag{10c}$$

where  $BER_i$  is the BER for UE*i*. This problem will give the rates for both UEs, and the optimization order, i.e., swapping indices 1 and 2, will not affect the achievable rate region.

The throughput  $\hat{R}_i$  in megabits per second [Mbps] and the spectral efficiency  $\hat{r}_i$  for user  $i$  have the following relation

$$\hat{R}_i = \hat{r}_i B. \tag{11}$$

Usually, coding efficiency  $\eta_i = \rho_i \log_2 M_i$  for UE*i* is adjusted based on MCS. Thus, if we can adjust  $k_i$ , which is equivalent to vary the symbol rate  $R_s$ , the solution space of the achievable rates can be enlarged.

It is worth mentioning that employing varying symbol rates does not mean AMC cannot be applied. On the contrary, the varying symbol rate scheme can help real-world AMC-based NOMA to reach a required throughput. The role of AMC is to provide a reasonable MCS corresponding to the channel gain.

### C. THE ALGORITHM FOR THE RATE REGION

Given the SNR of each user, one can solve (10) by an exhaustive search over  $C_3$ , i.e., by checking the feasibility of every candidate rate pair in  $C_3$ . However, traversing the entire code table is time-consuming. For this reason, we first find OMA points in Algorithm 1 and then use it to find NOMA rate pairs in Algorithm 2.<sup>1</sup> Algorithm 1 is to find the transmission rates of UE1 and UE2 achieved by OMA. First, the threshold  $\epsilon$  in (10c), modulation types, coding rates, and samples per symbol  $k$  can be set based on the requirement. Second, we decreasingly sort the code table based on the last column  $\tilde{r}$ . This step is a preparation for a bisection search. For each user  $i$ , using a bisection search over the sorted  $C_3$ , we search for the  $l_i^*$ th combination that satisfies the BER constraint, and record its corresponding combination of  $\{M, \rho, k, \tilde{r}\}$ . The order of UEs does not affect the results. For simplicity, we set UE1 first in Line 4. The outputs are the indices of rates  $\tilde{r}$  for UE1 and UE2 in the decreasingly sorted  $C_3$ .

---

#### Algorithm 1 Searching for OMA Rate pairs

---

- 1: Set a threshold  $\epsilon$  for BER;
  - 2: Set desired parameters for  $\{M, \rho, k\}$  and build  $C_3$ ;
  - 3: Sort  $C_3$  based on  $\tilde{r}$ 's column in descending order;
  - 4: **for** UE*i*  $i \in \{1, 2\}$  **do**
  - 5: Bisection search for the minimum index  $l_i^*$  in  $\{1, 2, \dots, |C_3|\}$  that ensures the  $BER_i \leq \epsilon$ ;
  - 6:  $l_i^{OMA} := l_i^*$ ;
  - 7: **end for**
  - 8: Output:  $l_1^{OMA}$  and  $l_2^{OMA}$ .
- 

Algorithm 2 starts based on the knowledge of OMA points obtained in Algorithm 1. We traverse the rest of the code table for UE2. This is because NOMA points on the rate region cannot exceed OMA boundary points. For each iteration,

<sup>1</sup>In the downlink NOMA, stronger users need to apply SIC. Thus, the algorithm complexity rises with the number of users. A practical way to limit complexity is to divide the users into groups. A Tx can exploit the spatial deployment to pair the users located close to the Tx and those near the cell edge. Within each group, SC and SIC are performed between a few users with disparate channels.



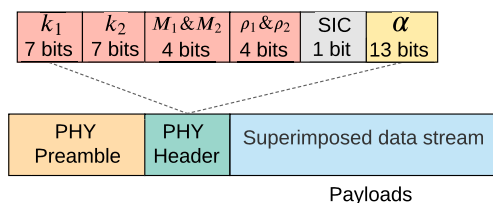
**Algorithm 2** Searching for NOMA Rate pairs

- 1: Set a threshold  $\epsilon$  for BER;
- 2: Set desired parameters for  $\{M, \rho, k\}$  and build  $C_3$ ;
- 3: Sort  $C_3$  based on  $\tilde{r}$ 's column in descending order;
- 4: Initialize  $\alpha := 1$ , and counter  $j := 0$ ;
- 5: **for** index  $l_2$  in  $\{l_2^{OMA}, \dots, |C_3|\}$  **do**
- 6:   Set the transmission parameters of UE2 same as those in the  $l_2$ th row in decreasingly sorted  $C_3$ ;
- 7:   Bisection search for maximum  $\alpha^* \in [0, \alpha]$  that ensures the  $BER_2 \leq \epsilon$ ;
- 8:   Bisection search for minimum  $l_1^*$  in  $\{l_1^{OMA}, \dots, |C_3|\}$  that ensures the  $BER_1 \leq \epsilon$ ;
- 9:   **if**  $l_1^*$  exists **then**
- 10:      $j := j + 1$ ;
- 11:     Record the  $j$ th iteration:  $\alpha := \alpha^*$ ,  $l_1^{NOMA} = l_1^*$ ,  $l_2^{NOMA} = l_2$ ;
- 12:   **else**
- 13:     Break;
- 14:   **end if**
- 15: **end for**
- 16: **Output:** A set of  $j$  feasible NOMA rate pairs with their  $\alpha^*$ ,  $l_1^{NOMA}$ , and  $l_2^{NOMA}$ .

the parameters of UE2 is first set according to row  $l_2$  in the sorted  $C_3$ . Then in Line 7, the power assigned to UE2 is expected to be less with the constraint  $BER_2$ , thus, UE1 can get more power (a higher  $\alpha$ ) to reach a higher rate. Next in Line 8, a bisection search is used to find the higher rate for UE1 that just satisfies  $BER_1$  and record the  $l_1^*$ th combination. Note that in the decreasingly sorted code table a smaller index  $l_1$  means a higher rate. The method for NOMA utilizes the test results of OMA and reduces the searching space and time compared with the Monte Carlo simulation which traverses possible combinations in code tables [9].

**D. FRAME STRUCTURE**

Transmission occurs in frames. The frame structure includes a preamble, header, and payloads as shown in Fig. 5. The preamble part is designed for data synchronization and channel equalization. It is a bit sequence that receivers start to locate the transmission. The header provides information about the packet configuration, such as MCS schemes, symbol rates, and power allocation coefficient. Finally, the payloads contain the data streams of UE1 and UE2.



**FIGURE 5.** The proposed frame structure for NOMA transmission.

1) SYNCHRONIZATION

Synchronization is one of the key procedures in practical communications. Accurate control of timing and the ability to synchronize operations will help improve the performance of SIC. Time synchronization is needed for symbol detection. We apply a predetermined threshold for rising edge detection. The decision of symbol timing is made according to the threshold calculated by the statistical properties of the noise.

We use trigger operation in the USRP transceiver to successfully collect pieces of complete and finite packets. Transmission in Tx is triggered once the host sends a “start trigger” command after running the code. In the two UE chains, the transceiver chooses the same trigger time as the Tx chain for timing. To make the transmission more practical, an idle gap is appended at the beginning of each transmission. The duration of the idle gap is unknown to the UEs. This simulates a practical scenario that UEs need to synchronize and detect the start of signals. Besides, due to hardware response delay, trigger operation alone cannot perform perfect frame synchronization. The threshold detection can be robust to detect the signal with delay and idle gap.

2) CHANNEL EQUALIZATION

The goal of equalization is to correct the distortion incurred by the channel and mitigate the effects of inter-symbol interference (ISI). We apply the zero-forcing (ZF) equalizer. Specifically, the ZF equalizer is designed as the inverse of the frequency response of the channel [55]. In the system design, the Tx sends the reference signals for channel estimation. Upon the reception of the pilots, the users perform channel equalization and calibrate the channel distortion.

3) NOMA HEADER DESIGN

We design the PHY layer header with the information that the proposed NOMA system requires. Specifically, in Fig. 5,  $k_1$  and  $k_2$  are 7-bit float binary numbers representing even decimal numbers in the range of 10 to 264.<sup>2</sup> The modulation types for OMA and the combined modulation types for NOMA include 15 possible cases, thus 4 bits are needed. Three coding rates  $\rho = \{1, \frac{1}{2}, \frac{3}{4}\}$  for each user are considered, where 1 means disabling the convolutional coding. Doing SIC is an on-off switch requiring one bit, “0” means to perform SIC, and “1” means to treat interference as noise. Finally, the power allocation factor  $\alpha$  supports four decimal digits which requires at least 13 bits. Based on IEEE 802.11b packet format [38], the modulation type for preamble and header is BPSK. As a low-rate modulation, BPSK can assure the important information is transmitted correctly.

Besides, for implementation simplification, in [9], [16], it is assumed that the receivers know the coding and power allocation information, which are not practical assumptions.

<sup>2</sup>In this system design, the transceiver has the same sampling rate for Tx and users. If different sampling rates are considered, then a symbol rate or symbol duration should be transmitted.

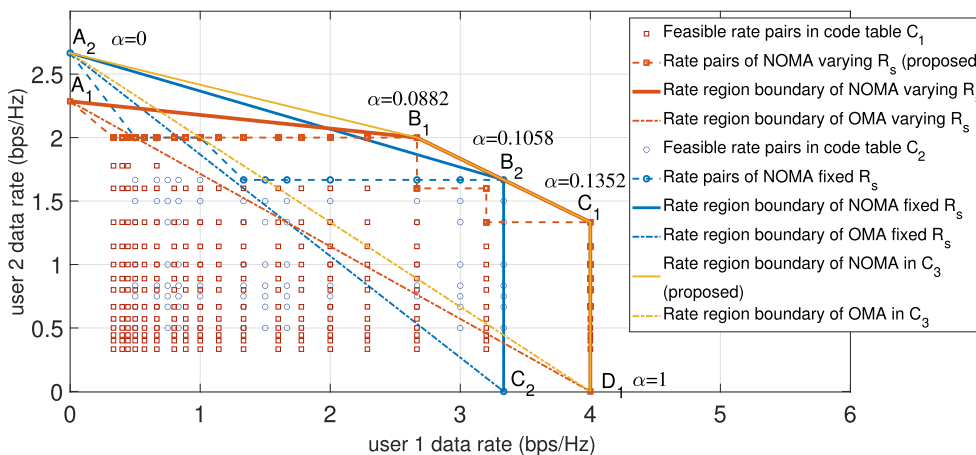


FIGURE 6. Comparison on varying and fixed symbol rates in a NOMA transmission with  $\gamma_1 = 17.7\text{dB}$ ,  $\gamma_2 = 5.7\text{dB}$ .

In our experiment, we transmit NOMA parameters over the air.

VI. SIMULATION RESULTS AND ANALYSIS

Before we move to the implementation results, we perform simulation without imperfectness, i.e., we assume synchronization is perfect, and channel and NOMA header are known. The goal is to validate the varying symbol rate method on MATLAB R2020a. In this simulation, we include three scenarios:

- Varying symbol rate, which results in varying samples per symbol  $k_i = \{8, 10, \dots, 24\}$ , and  $\rho = 1$ .
- Fixed symbol rate ( $k_i = 8$  for each user) and varying coding rate  $\rho = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\}$ .
- Varying symbol rate with  $k_i = \{8, 10, \dots, 24\}$ , and  $\rho = \{1, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\}$ .

In all cases, the sampling rate  $f_s$  is 10 Msps, and modulation types include BPSK, QPSK, and 16QAM. For the first case (proposed), the feasible solutions are from code table  $C_1$ . We vary the samples per symbol  $k_i$  from 8 to 24 with a step two in which the symbol rate varies correspondingly from  $4.167 \times 10^5$  to  $1.8 \times 10^6$  symbols per second. In the second case, the MCS with coding rate  $\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\}$  [9] is from code table  $C_2$ . In the third case (proposed), feasible solutions are from code table  $C_3$  which is to search over  $\rho = \{1, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}\}$ ,  $k_i = \{8, 10, \dots, 24\}$ , and  $k_{\min} = 8$  in this simulation.

The results are shown in Fig. 6, where  $\gamma_i$  as the estimated SNR at  $UE_i$  which is measured by

$$\gamma_i = \frac{P_i - P_{N_0}}{P_{N_0}} = \frac{\frac{1}{N_s} \sum_{j=1}^{N_s} y_i(j)y_i^*(j) - \frac{1}{n} \sum_{k=1}^n |n_i(k)|^2}{\frac{1}{n} \sum_{k=1}^n |n_i(k)|^2}, \quad (12)$$

in which  $N_s$  is the number of samples for  $UE_i$ ,  $P_i$  is the received signal power at  $UE_i$ ,  $P_{N_0}$  is the noise power at

antennas of UEs, and  $n$  is the number of noise samples (the noise power will be measured separately when there is no signal). For all cases, the SNRs are set as  $\gamma_1 = 17.7\text{dB}$  and  $\gamma_2 = 5.7\text{dB}$ . For a fair comparison with the second case in [9], where the unit is bits per second per Hz (bps/Hz) with packet error rate (PER),<sup>3</sup> we set  $\epsilon = 10^{-4}$ . The feasible rate pairs are limited within  $\epsilon$ . The red square points denote the feasible rate pairs in  $C_1$  with varying  $R_s$ . The red dash line is the optimal rate-pairs obtained by solving (10). The rate region boundary is a convex hull of the optimal set along with the single-user OMA points. The fixed symbol rate cases with MCS are denoted by blue color. The yellow curve denotes the rate region obtained from  $C_3$ .

For UE1 OMA rate point  $D_1$  achieved by varying  $R_s$  is larger than point  $C_2$  achieved by fixed  $R_s$ . This is because UE1 has a better channel and thus it can reach the system limit, i.e.,  $\rho = 1$  for 16QAM, and  $R_s = 1.8 \times 10^6$  with  $k_1 = 8$ . Point  $C_2$  is reached by setting  $\rho = 5/6$  for 16QAM. On the other hand, point  $A_2$  ( $\rho = 2/3$  for 16QAM, and  $R_s = 1.8 \times 10^6$  with  $k_1 = 8$ ) is greater than point  $A_1$  ( $\rho = 1$  for 16QAM, and  $R_s = 7.14 \times 10^5$  with  $k_1 = 14$ ). The reason is that UE2 suffers from a poor channel thus channel coding is required to reach a high rate while satisfying  $BER < \epsilon$ . In the achievable NOMA rate pairs, points  $B_1$  and  $C_1$  span a larger rate region ( $A_1 \rightarrow B_1 \rightarrow C_1 \rightarrow D_1$ ) compared with the fixed  $R_s$  ( $A_2 \rightarrow B_2 \rightarrow C_2$ ). The combination of the two code tables, or say, the existing MCS with varying symbol rates, can further enlarge the rate region ( $A_2 \rightarrow B_1 \rightarrow B_2 \rightarrow C_1 \rightarrow D_1$ ) for NOMA compared with applying only MCS [9]. Table 7 shows the detailed configuration of the NOMA rate pairs.

An observation from Fig. 6 is that the rate pairs below the OMA rate region do not have to be searched. Thus, we can apply a Monte Carlo simulation via Algorithm 1 to first find the OMA points to reduce the search space by removing the

<sup>3</sup>PER in [9] is 0.1, and  $PER \approx BER \times n$ , where  $n$  is the length of the packet in bits.

TABLE 7. Simulation parameters for each point in Fig. 6.

Simulation: $\gamma_1 = 17.7\text{dB}$ , $\gamma_2 = 5.7\text{dB}$ .							
Point	$\alpha$	$M$		$\rho$		$k$	
		UE1	UE2	UE1	UE2	$k_1$	$k_2$
A <sub>1</sub>	0	-	16	-	1	-	14
B <sub>1</sub>	0.0882	16	4	1	1	12	8
C <sub>1</sub>	0.1352	16	4	1	1	8	12
D <sub>1</sub>	1	16	-	1	-	8	-
A <sub>2</sub>	0	-	16	-	2/3	-	8
B <sub>2</sub>	0.1058	16	4	5/6	5/6	8	8
C <sub>2</sub>	1	16	-	5/6	-	8	-

TABLE 8. Average execution time comparison over different methods.

Method	Time
Monte Carlo simulation [9]	4 hours
Bisection search over C <sub>1</sub>	4.08min
Bisection search over C <sub>2</sub>	2.85min
Bisection search over C <sub>3</sub>	47.52min

rate pairs bellow the OMA rate region. The operation time is listed in Table 8. The Monte Carlo simulation realized in [9] takes about 4 hours. Reproducing their main work (without OFDM) taking about one and a half hours in our simulations. We apply bisection search over C<sub>2</sub> which reduces the operation time from hours to several minutes. Besides, using C<sub>3</sub> takes more time since the number of feasible rate pairs is increased. We ran simulations on a Windows 10 workstation with CPU 3.2GHz and with 16GB RAM, compared with dual-core Linux running at 2.4GHz with 2GB RAM in [9].

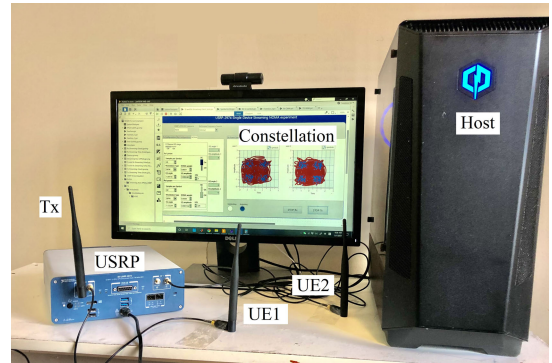
## VII. EXPERIMENTAL RESULTS AND ANALYSIS

### A. EXPERIMENT SETUP

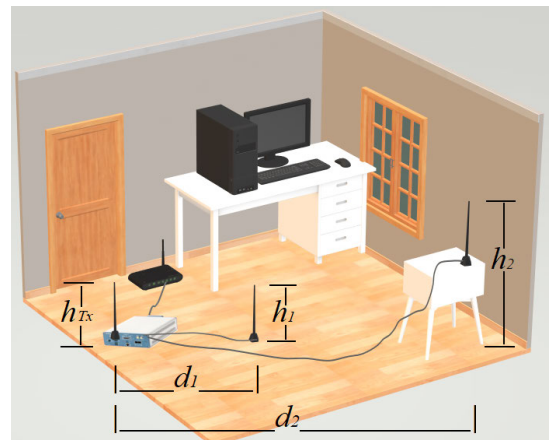
#### 1) THE TESTBED

We experiment the two-user SISO system over one SDR transceiver USRP 2974 and a Window 10 system host whose central processing unit (CPU) model is AMD Ryzen 3 2300X. We develop NOMA over a USRP 2974 single-device streaming sample project software [13] on LabVIEW communications system design suite [14], which is supported by the LabVIEW NXG environment. The USRP 2974 transceiver supports the inner loop mode allowing signal transmission and reception. The host and USRP device should be in the same wireless local area network (WLAN). We connect USRP and the router with one Ethernet cable. USRP 2974 can support a frequency range from 10 MHz to 6 GHz. In the experiment, the carrier frequency is set as 3GHz. Based on the latest release of 3GPP, we are sitting on the frequency range 1 (FR1) of 5G new RF bands, which is from 410 MHz to 7125 MHz [56].

The experimental environment and the space map of our test scenarios are shown in Fig. 7. Regarding the environments, the Tx is placed in an indoor environment with a height of 23cm. UE1 and UE2 are placed at distances  $d_1$  and  $d_2$  from



(a) Experimental environment



(b) Testbed space map

FIGURE 7. NOMA system experimental environment and space map.

TABLE 9. Parameters of the experiment.

Parameter	Parameter Description	Value
$f_c$	Carrier frequency	3GHz
$B$	Bandwidth	1MHz
$P$	Transmitter power	-10 ~ 20 dBm
$N_s$	Number of samples for Tx/UEs	21160/30000
$k_i$	Samples per symbol	{10, 12, ..., 128}
$N_{sy}$	Number of symbols	$N_s/k$
$f_s$	Sampling rate	10 mega samples/s
$R_s$	Symbol rate	$f_s/k$
$M$	Modulation type for UEs	$M = \{2, 4, 16\}$
$\rho$	Coding rate	$1, \frac{1}{2}, \frac{3}{4}$

the Tx with heights  $h_1$  and  $h_2$ .<sup>4</sup> The fundamental experiment parameters and configurations for Tx and UEs are denoted in Table 9.

#### 2) POWER CONTROL

We adjust the SNR for each user according to antenna direction and locations. The work in [9] has emulated a perfect power-controlled environment by connecting the Tx and UEs with coaxial cables to mitigate the propagation loss caused by

<sup>4</sup>Changing the direction of antennas or user locations will change the SNR. In this paper, we slightly adjust the direction of the Tx antenna to reach the desired SNR within a limited space.

physical environment variation. In this paper, the experiment is carried out in an indoor wireless environment within a stable temperature.

**B. EXPERIMENT SCENARIOS**

We consider three scenarios in NOMA, which are

- $\gamma_1 \gg \gamma_2$  if  $\gamma_1 - \gamma_2 > 10\text{dB}$ ;
- $\gamma_1 > \gamma_2$  if  $10\text{dB} \geq \gamma_1 - \gamma_2 > 3\text{dB}$ ;
- $\gamma_1 \approx \gamma_2$  if  $3\text{dB} \geq \gamma_1 - \gamma_2 \geq 0\text{dB}$ .

*Example 1* ( $\gamma_1 \gg \gamma_2$ ):

The measured SNRs for user 1 and user 2 are  $\gamma_1 = 17.7\text{dB}$  and  $\gamma_2 = 5.7\text{dB}$  according to (12). The distances and heights of each user in Fig. 7(a) for the SNR configuration is shown in Table 10-Example 1.

**TABLE 10. Parameter settings of the achievable rates.**

Example 1: $\gamma_1 = 17.7\text{dB}$ , $\gamma_2 = 5.7\text{dB}$ .					
$d_1$	$d_2$	$h_1$	$h_2$	$h_{Tx}$	$P$
60cm	352cm	20cm	277cm	21.8cm	10dBm
$\alpha$	$k_1$	$k_2$	$M_1$	$M_2$	$(R_1, R_2)$
0	-	16	-	4	(0, 1.43)
0.0301	18	16	4	4	(1.11, 1.25)
0.0386	16	18	4	4	(1.25, 1.11)
0.0631	16	20	4	4	(1.25, 1.00)
0.1259	16	30	4	4	(1.25, 0.67)
1	22	-	16	-	(1.82, 0)

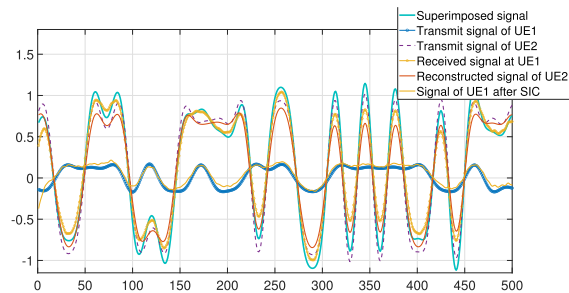
  

Example 2: $\gamma_1 = 18.4\text{dB}$ , $\gamma_2 = 10.5\text{dB}$ .					
$d_1$	$d_2$	$h_1$	$h_2$	$h_{Tx}$	$P$
60cm	314cm	20cm	80cm	23cm	10dBm
$\alpha$	$k_1$	$k_2$	$M_1$	$M_2$	$(R_1, R_2)$
0	-	24	-	16	(0, 1.67)
0.08	28	18	4	4	(0.71, 1.11)
0.10	18	18	4 ( $\rho = \frac{3}{4}$ )	4	(0.83, 1.11)
0.10	26	18	4	4	(0.77, 1.11)
1	22	-	16	-	(1.82, 0)

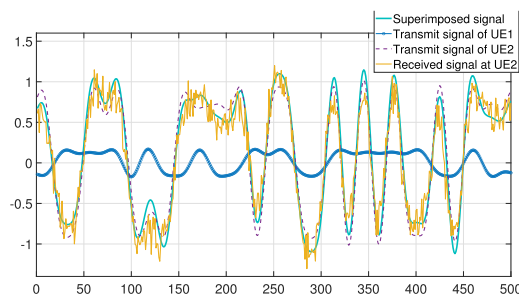
  

Example 3: $\gamma_1 = 8.8\text{dB}$ , $\gamma_2 = 7.3\text{dB}$ .					
$d_1$	$d_2$	$h_1$	$h_2$	$h_{Tx}$	$P$
cm	cm	80cm	277cm	23cm	0dBm
$\alpha$	$k_1$	$k_2$	$M_1$	$M_2$	$(R_1, R_2)$
0	-	12	-	4	(7, 1.67)
0.0308	60	16	4	4	(0.33, 1.25)
0.05	52	18	4	4	(0.38, 1.11)
1	12	-	4	-	(1.67, 0)

The encoded signals with power allocation of UE1 (blue circle line) and UE2 (purple dash line) are shown in Fig. 8 and Fig. 9. The superimposed signal filtered by root raised cosine filter with roll off factor  $\beta = 0.5$  is denoted as the green solid line. At UE1, the received in-phase (I) link signal after equalization with power allocation coefficient  $\alpha = 0.0301$  is shown in Fig. 8 in yellow circle line (only 500 points are plotted). We demodulate and decode from the received signal at UE1 (yellow circle line in Fig. 8), and then reconstruct the signal for UE2 with the knowledge of UE2's modulation type and power allocation coefficient. The value of  $\alpha$  affects the amplitude of the reconstructed signal. The red line represents the reconstructed signal of UE2 at UE1. Finally, the signal of UE1 (yellow solid line) is obtained by subtracting the UE2's signal from the received signal at UE1, i.e., the yellow circle line minus the red line. Figure 9 shows the received signal



**FIGURE 8. Transmitted and received signals at UE1 before and after SIC with  $\alpha = 0.0301$  and  $\gamma_1 = 17.7\text{dB}$ .**



**FIGURE 9. Transmitted and received signals at UE2 with  $\gamma_2 = 5.7\text{dB}$ .**

at UE2. The blue line denotes the I link while the purple dashed line represents the quadrature (Q) link. In Fig. 8 and Fig. 9, both blue lines reflect the superimposed signals while suffering different channel impairments.

Figure 10(a) shows the rate pairs. The OMA rate region (yellow line) connecting the two marginal points is obtained by time-sharing. The circle points represent the measured rate pairs obtained by solving (10). To obtain the rate region boundary, we find the convex hull of the set including single-user OMA points to have a convex polygon that is in a solid blue line.

To obtain the achievable rate regions in Fig. 10(a), we can adjust the power allocation coefficient. When  $\alpha = 1$ , the P2P communication of UE1 is realized, which is also the point  $(R_1, 0)$  in the rate region boundary for OMA and an extreme case of NOMA. To reach this point, the modulation type is set as 16QAM ( $M = 16$ ), and the minimum  $k_s$  is measured at 22 while keeping the threshold  $\epsilon$  in (10) of BER less than 0.1%. Thus, the rate obtained from (11) is  $\hat{R}_1 = \log_2(M_1) \frac{f_s}{k_1} = \log_2 16 \cdot \frac{10 \cdot 10^6}{22} = 1.82 \text{ Mbps}$ . For the rate pairs  $(R_1, R_2)$  achieved by SC and SIC, the modulation types and symbol rates for both users can be measured by Algorithm 2. For example, when  $\alpha = 0.0301$ , UE1 and UE2 are both QPSK, i.e.,  $M_1 = M_2 = 4$ , and the samples per symbol for UE1  $k_1$  is 18 and  $k_2$  is 16. The decoded QPSK constellation for each user is shown in Fig. 11(b). Other values of  $\alpha$  reaching the boundary in Fig. 10(a) are shown in Table 10. The constellation for the boundary points is shown in Fig. 11(b), in which the red points are the samples, and the blue points are the decoded symbols.

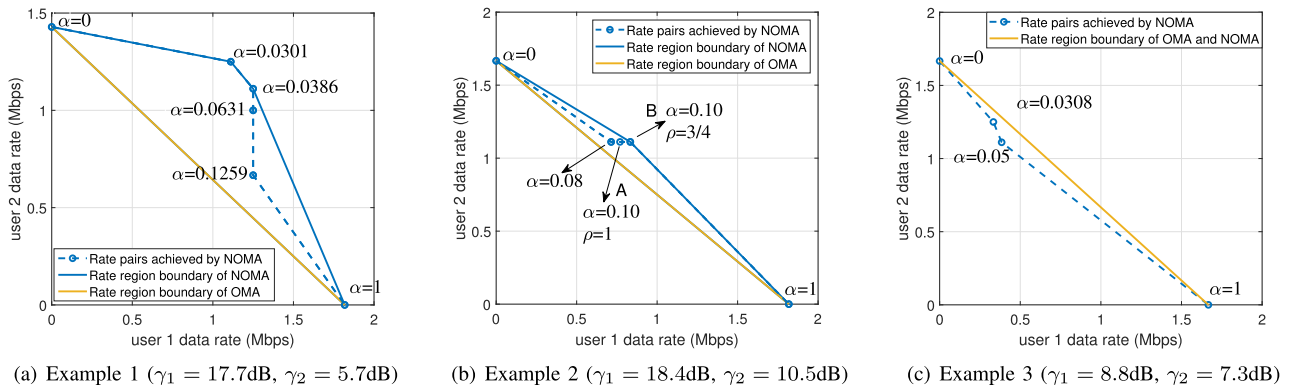


FIGURE 10. Experimentally obtained rate pairs for Example 1, Example 2, and Example 3.

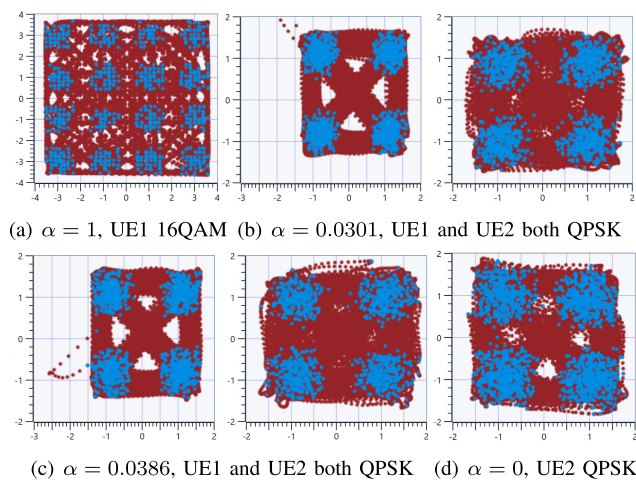


FIGURE 11. Constellation of Example 1.

Example 2 ( $\gamma_1 > \gamma_2$ ):

In this case,  $\gamma_1 = 18.4\text{dB}$  and  $\gamma_2 = 10.5\text{dB}$ . The configuration of each user is shown in Table 10-Example 2. Figure 10(b) shows the rate pairs. The rate region is getting close to the OMA scheme because the difference between channel gains has reduced. The measurement of the blue points in Fig. 10(b) is very similar to that in Example 1. Two points in Fig. 10(b) are very interesting, *point A* ( $\rho = 1$ ) and *point B* ( $\rho = \frac{3}{4}$ ) with the same  $\alpha = 0.1$  and modulation type. The *point B* with  $\rho = \frac{3}{4}$  can support a higher symbol rate, which can compensate for the cost of coding rate reduction in this experiment. In fact, without convolutional coding, the minimum  $k_1$  for an acceptable BER is 28, whereas with  $\rho = \frac{3}{4}$ ,  $k_1$  and  $k_2$  are 18 for both users, the transmission rate enabling convolutional coding with fixed symbol rate can be higher than that varying symbol rate. In the experiment, BER for *point A* is  $\epsilon = 0$ , while BER for *point B* is  $\epsilon \in [0\%, 0.18\%]$ .

We show the decoded constellation for each user in Fig. 12. Specifically, the two P2P points for UE1 and UE2 are shown in Fig. 12(a) and Fig. 12(b), respectively. The constellation for the boundary point  $\alpha = 0.1, \rho = \frac{3}{4}$  is shown in Fig. 12(c), in which a set of red samples are jumping away from the

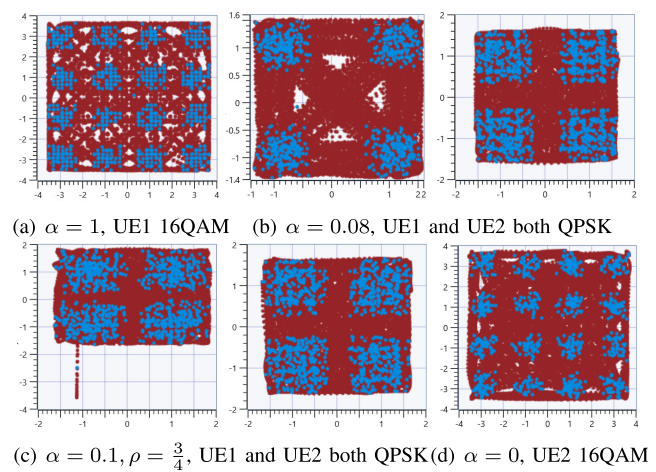


FIGURE 12. Constellation of Example 2.

constellation possibly due to the start-oscillation. The four clusters of QPSK get closer and blurrier with a higher rate pair (0.83, 1.11) compared with that of Fig. 12(b) where a lower rate pair (0.71, 1.11) is tested. This is reasonable because, with a lower rate, the decoding performance is better, thus, the constellation in Fig. 12(b) is more distinguishable than Fig. 12(c).

Example 3 ( $\gamma_1 \approx \gamma_2$ ):

When  $\gamma_1 = \gamma_2$ , the channel is called symmetric. The capacity region for OMA and NOMA should be identical [57]. In this experiment, we set  $\gamma_1 = 8.8\text{dB}$  and  $\gamma_2 = 7.3\text{dB}$ . From the results in Fig. 10(c) and Fig. 13, it is hard to find a NOMA rate pair lying on the rate region. In other words, the point lying on the rate region in practice will have an unacceptable BER.

In summary, when  $\gamma_1$  is much higher than  $\gamma_2$  in Example 1, SIC can be implemented successfully with a low BER. NOMA points can be easily found and can outperform OMA rates. We also find that in some cases, such as Example 2, MCS along with convolutional coding is very useful to reduce the BER and at the same time maintain a high rate. When the channel gains become closer, the NOMA rate region could be even lower than that of OMA, which is the reason why

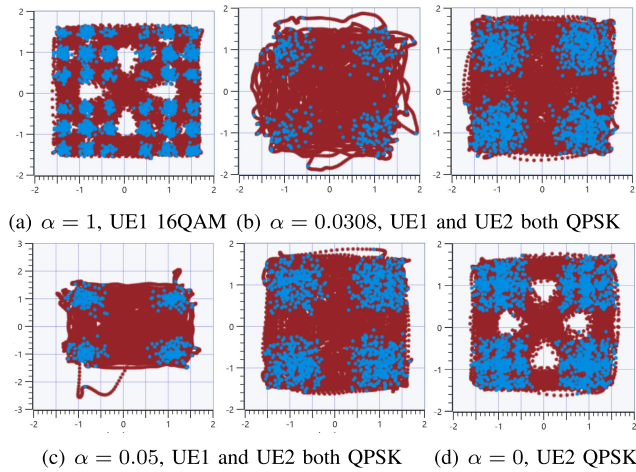


FIGURE 13. Constellation of Example 3.

many papers have the assumption  $|h_1|^2 > |h_2|^2$  to ensure a successfully encoding and decoding in NOMA.

C. GAP BETWEEN PRACTICE AND THEORY

In this part, we discuss the gap between simulation, experiments, and capacity. To illustrate this, let  $\gamma_1 = 9.5\text{dB}$  and  $\gamma_2 = 4\text{dB}$ . The theoretical SNR is defined as

$$\gamma_1^{th} = \frac{P|h_1|^2}{N_0} \text{ and } \gamma_2^{th} = \frac{P|h_2|^2}{N_0}, \quad (13)$$

where  $N_0$  is the power spectral density of the noise. The noise power in the measured  $\gamma_1$  and  $\gamma_2$  is over the whole bandwidth  $W = f_s$ . The relation between  $\gamma_i^{th}$  and measured  $\gamma_i$  given in (12) is

$$\gamma_i^{th} = \frac{P_i - P_{N_0}}{\frac{P_{N_0}}{W} \cdot B} = \frac{P_i - P_{N_0}}{P_{N_0}} k_{\min} = \gamma_i \cdot k_{\min}. \quad (14)$$

The experimental rates are (11) and (10) while the theoretical capacity region is obtained in (5). Similar to Figs. 10(a)-10(c), in Fig. 14, the yellow line shows the OMA rate region obtained by time-sharing the experimental results. The blue circle points denote the measured rate pairs obtained by Algorithm 2. The solid blue line represents the convex hull of all measured rate pairs. The green line is achieved using code table  $C_3$  in Section VI with  $k_{\min} = 10$ . Each green circle denotes a rate pair obtained by simulation. The simulation is the ideal case of implementation in which practical issues such as synchronization error, channel estimation errors, SIC error are not there. Those issues and hardware limits reduce the transmission rate in practice.

The theoretical capacity is shown by a red line. Points  $A_1, B_1, C_1, D_1, E_1,$  and  $F_1$  on the boundary of the NOMA capacity region are obtained for  $\alpha$  at 0, 0.05, 0.25, 0.5, 0.8, and 1, correspondingly. The experimental values of  $\alpha$  at corner points  $A_2, B_2, C_2,$  and  $D_2$  are 0, 0.1081, 0.1918, and 1, respectively. In theory, to achieve the capacity, the power allocation  $\alpha$  could be any number between 0 and 1 [41, Chapter 6.2.2, pp. 279], i.e., the power allocated to the

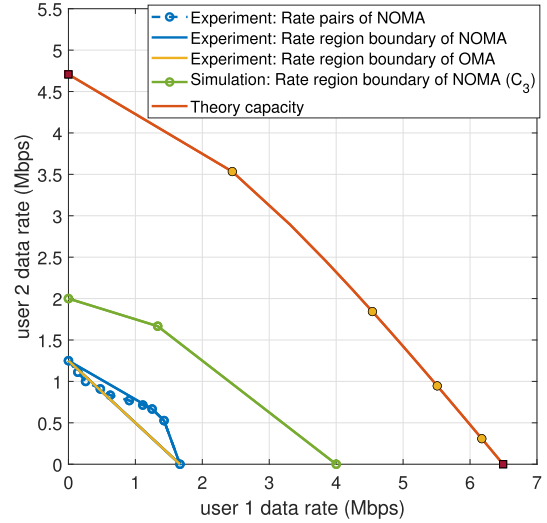


FIGURE 14. A comparison among experimental, simulation-based, and theoretical achievable regions.

weak user (UE2) can be higher than, equal to, or less than that of the other user. That is,  $|h_1| > |h_2|$  does not imply  $\alpha < 1/2$  and it is not necessary to allocate less power to the strong user [45, Myth 1]. In a practical communication system where a specific modulation type is used, the power allocation may have certain constraints to avoid constellation overlap and successful decoding [43], [44].

The gap between practical experiments and theoretical capacity is still noticeable. Although AMC is well-developed and favored as a baseline scheme in practice, it is sub-optimal and cannot achieve the Shannon capacity. As seen in [25, Fig. 3] and [58, Fig. 5], there is a significant gap between Shannon capacity and the data rate achieved by the MCS-based approach. Specifically, Shannon capacity [59] is based on the assumption that a code table with infinite length and the elements therein drawn from a Gaussian distribution while the inputs of MCS are discrete and have a finite length. Therefore, the system performance will not achieve capacity with a near-zero BER.

VIII. FUTURE DIRECTIONS

We describe some important practical issues as well as the future directions for NOMA experiments in this section.

Several experiments need to be carried out before NOMA being accepted in wireless communications standards. We discuss some future directions here.

A. MIMO-NOMA

As listed in Table 1, SISO-NOMA has been experimented with by various groups whereas the number of experiments on MIMO-NOMA is limited. More importantly, those results are using SC-SIC for MIMO-NOMA while SC-SIC is not optimal in multi-antenna BC due to the non-degradedness of BC [45], [60]. It is well-known in information theory that dirty paper coding (DPC) is the capacity-achieving scheme for the multiple antenna BC [45], [60]–[62]. Due to the high

computational cost of DPC, linear precoding is used for the transmit signal design [63]. Linear precoding can achieve the same region as DPC achieves. MIMO-NOMA [19], [20] and MISO-NOMA [25] require precoding and power allocation to achieve spatial multiplexing, thus, the precoders are sent in the preamble before the data transmission. Low-complexity linear precoding schemes to reach the capacity, such as iterative water-filling [64], simultaneous triangularization [65] can be adjusted to practical constraints and applied in future practical experiments.

### B. MULTI-CELL NOMA

All existing experimental results are based on one Tx, i.e., a single-cell network. Today's cellular networks are multi-cell, however. In a multi-cell network, inter-cell interference (ICI) is a major issue that affects the performance of cell-edge users. A classical method to avoid interference at receivers is to orthogonalize the transmissions of different users, which is not spectral efficient. In single antenna networks, Han-Kobayashi (HK) encoding the best-known scheme [66]. HK is based on decoding part of interference and treating part of that as noise. In MIMO systems, interference alignment (IA) is a cooperative transmission technique to remove the ICI in which the Tx's jointly optimize their beamforming vectors. IA validation via testbed measurements can be found in [67]–[72] and the references therein.

The above IA implementations are based on OMA. The combination of IA with NOMA increases the number of users and spectral efficiency in multi-cell systems [73], [74], and is worth implementation trials if NOMA is to welcome to practical networks. To perform multi-cell NOMA, normally, global CSI is required to be perfectly known at all terminals. However, acquiring such channel knowledge is a challenging problem in practice. Blind interference alignment techniques are important to this end [75].

### C. SYSTEM-LEVEL IMPLEMENTATION

A system-level implementation refers to the large-scale deployment of NOMA with multiple Tx's and Rx's in multi-cells. In the large-scale NOMA, two key issues may affect the performance. First, the high inter-user-interference, which is the interference caused by other NOMA users sharing the same resource [76]. Second, the high hardware complexity requirements of SIC scales significantly with the number of NOMA users [73], [76]. To solve these issues, coordinated and joint transmission NOMA schemes pair the cell-edge users and cell-center users as a NOMA group, and then reduce the inter-user-interference and enhance the diversity [76]. In an implementation, synchronization should be considered and, PSS and SSS will be applied for cell search.

### D. SIGNAL PROCESSING USING MACHINE LEARNING

Machine learning (ML) and deep learning are becoming a ubiquitous solution in wireless communications, e.g., in blind detection of modulation orders [77], modulation classification [78], [79], real-time interference cancellation [80],

precoding design in MIMO networks [81], and so on. ML could be used in real-time SDR-based experiments [80] or for offline processing. Real-time application in the context of NOMA could be SIC and synchronization. Possible applications of ML in offline processing could be in symbol detection and BER reduction. ML also can be used to design a fast and efficient precoder [81] in MIMO-NOMA experiments. This is a processing field and is expected to flourish in near future.

### IX. CONCLUSION

We have developed a NOMA wireless transmission system on LabVIEW NXG and USRP. We implemented the key NOMA methods, SC at the Tx and SIC at the strong user side while the other user treats interference as noise. Our designed system can support a varying transmission symbol rate with and without convolutional coding. By varying the symbol rate of each NOMA user, the code table or solution space based on existing MCS can be enlarged which, in turn, can provide additional rate-achieving schemes in terms of modulation types, coding rate, and symbol rates. This variation can help NOMA users to achieve a higher rate region within acceptable BERs. We propose two algorithms to efficiently find the achievable rate pairs for OMA and NOMA. The algorithm is time-saving compared with Monte Carlo simulation over all possible rate pairs combinations. We have also provided a detailed design for the NOMA protocol including preamble and NOMA headers. The results verify that NOMA can achieve a better rate region than OMA in asymmetric channels, but there still exists space for practical NOMA implemented by AMC to further bridge the gap between the theory and experiment. Finally, we have discussed implementation challenges and some possible future directions.

### ACKNOWLEDGMENT

The authors would like to thank NI for their technical guidance in hardware issues and software support. They also thank the Associate Editor and anonymous reviewers for their valuable suggestions during the editorial process.

### REFERENCES

- [1] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Cham, Switzerland: Springer, 2019.
- [2] A. Benjebbour, A. Li, K. Saito, Y. Saito, Y. Kishiyama, and T. Nakamura, "NOMA: From concept to standardization," in *Proc. IEEE Conf. Standards for Commun. Netw. (CSCN)*, Oct. 2015, pp. 18–23.
- [3] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 611–615.
- [4] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.
- [5] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.
- [6] M. Vaezi, G. A. Aruma Baduge, Y. Liu, A. Arafa, F. Fang, and Z. Ding, "Interplay between NOMA and other emerging technologies: A survey," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 4, pp. 900–919, Dec. 2019.

- [7] *Study on Downlink Multiuser Superposition Transmission*, document 3GPP TD RP-150496, 2015.
- [8] H. Lee, S. Kim, and J.-H. Lim, "Multiuser superposition transmission (MUST) for LTE-A systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
- [9] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2628–2639, Jul. 2012.
- [10] X. Xiong, W. Xiang, K. Zheng, H. Shen, and X. Wei, "An open source SDR-based NOMA system for 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 24–32, Dec. 2015.
- [11] X. Wei, H. Liu, Z. Geng, K. Zheng, R. Xu, Y. Liu, and P. Chen, "Software defined radio implementation of a non-orthogonal multiple access system towards 5G," *IEEE Access*, vol. 4, pp. 9604–9613, 2016.
- [12] J. Mitola, III, "Software radios: Survey, critical evaluation and future directions," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 8, no. 4, pp. 25–36, Apr. 1993.
- [13] National Instruments. *Specification USRP-2974 USRP Software Defined Radio Stand-Alone Device*. Accessed: Sep. 2021. [Online]. Available: <https://www.ettus.com/wp-content/uploads/2019/01/377417a.pdf>
- [14] (2019). *LabVIEW Communications System Design Suite 4.0 Manual*. [Online]. Available: <https://www.ni.com/documentation/en/labview-comms/4.0/manual/manual-overview/>
- [15] *NI USRP and LabVIEW Communications System Design Suite*. Accessed: Sep. 2021. [Online]. Available: <https://www.ettus.com/wp-content/uploads/2019/01/377417a.pdf>
- [16] Z. Gong, M. Haenggi, S. Srinivasa, K. Stamatiou, S. Vanka, and P. Vizi. *Technical Report for Superposition Coding Strategies: Design and Experimental Evaluation*. Accessed: Sep. 2021. [Online]. Available: [https://www3.nd.edu/~mhaenggi/pubs/techreport\\_sc11.pdf](https://www3.nd.edu/~mhaenggi/pubs/techreport_sc11.pdf)
- [17] *IEEE Standard for Information Technology: Telecommunications and Information Exchange Between System Local and Metropolitan Area Networks, Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard IEEE 802.11ax, 2019.
- [18] A. Benjebbour and Y. Kishiyama, "Combination of NOMA and MIMO: Concept and experimental trials," in *Proc. 25th Int. Conf. Telecommun. (ICT)*, Jun. 2018, pp. 433–438.
- [19] A. Benjebbour, Y. Kishiyama, Y. Okumura, C.-H. Hwang, and I.-K. Fu, "Outdoor experimental trials of advanced downlink NOMA using smartphone-sized devices," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 1–6.
- [20] A. Benjebbour, K. Saito, and Y. Kishiyama, "Experimental trials on non-orthogonal multiple access," in *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Cham, Switzerland: Springer, 2019, pp. 587–607.
- [21] X. Wei, Z. Geng, H. Liu, K. Zheng, and R. Xu, "A portable SDR non-orthogonal multiple access testbed for 5G networks," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–5.
- [22] H. Pan, L. Lu, and S. C. Liew, "Practical power-balanced non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2312–2327, Oct. 2017.
- [23] L. Y. Hosni, A. Y. Farid, A. A. Elsaadany, and M. A. Safwat, "5G new radio prototype implementation based on SDR," *Commun. Netw.*, vol. 12, no. 1, pp. 1–27, 2020.
- [24] S. Abeywickrama, L. Liu, Y. Chi, and C. Yuen, "Over-the-air implementation of uplink NOMA," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.
- [25] 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.
- [26] J.-R. Garnier, A. Fabre, H. Fares, and R. Bonnefoi, "On the performance of QPSK modulation over downlink NOMA: From error probability derivation to SDR-based validation," *IEEE Access*, vol. 8, pp. 66495–66507, 2020.
- [27] M. A. Durmaz, E. Guven, G. K. Kurt, M. Toka, M. Aldababsa, and O. Kucur, "A four-user non-orthogonal multiple access system implementation in software defined radios," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, May 2020, pp. 1–5.
- [28] M. Aldababsa, C. Goztepe, G. K. Kurt, and O. Kucur, "Bit error rate for NOMA network," *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1188–1191, Jun. 2020.
- [29] E. Khorov, A. Kureev, and I. Levitsky, "NOMA testbed on Wi-Fi," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2018, pp. 1153–1154.
- [30] E. Khorov, A. Kureev, I. Levitsky, and I. F. Akyildiz, "Prototyping and experimental study of non-orthogonal multiple access in Wi-Fi networks," *IEEE Netw.*, vol. 34, no. 4, pp. 210–217, Jul. 2020.
- [31] A. A. Kureev, I. A. Levitsky, and E. M. Khorov, "Experimental study of constellation rotation in NOMA Wi-Fi networks," *J. Commun. Technol. Electron.*, vol. 65, no. 12, pp. 1525–1530, Dec. 2020.
- [32] B. S. K. Reddy, "Experimental validation of non-orthogonal multiple access (NOMA) technique using software defined radio," *Wireless Pers. Commun.*, vol. 116, pp. 3599–3612, Nov. 2020.
- [33] J. Angjo, M. M. Tuncer, E. Akertek, H. Alakoca, M. Basaran, and L. Durak-Ata, "On the channel estimation performance of NOMA systems: Experimental implementation of real-time downlink NOMA-OFDM," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, May 2020, pp. 1–6.
- [34] A. R. Khan and S. Sohaib, "Cooperative NOMA, prototyping and experimental evaluation using SDR," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2872–2876, Mar. 2021.
- [35] E. Khorov, A. Kureev, I. Levitsky, and I. F. Akyildiz, "Prototyping NOMA constellation rotation in Wi-Fi," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 1332–1333.
- [36] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613–1621, Dec. 1997.
- [37] C. Cox, *An Introduction to LTE: LTE-advanced, SAE and 4G Mobile Communications*. Hoboken, NJ, USA: Wiley, 2012.
- [38] Primer, *Wi-Fi: Overview of the 802.11 Physical Layer and Transmitter Measurements*. Tektronix, Beaverton, OR, USA, 2013, pp. 4–7.
- [39] F. Ling, *Synchronization in Digital Communication Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.
- [40] X. Lu, L. Ni, S. Jin, C.-K. Wen, and W.-J. Lu, "SDR implementation of a real-time testbed for future multi-antenna smartphone applications," *IEEE Access*, vol. 5, pp. 19761–19772, 2017.
- [41] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [42] A. Li, Y. Lan, X. Chen, and H. Jiang, "Non-orthogonal multiple access (NOMA) for future downlink radio access of 5G," *China Commun.*, vol. 12, pp. 28–37, Dec. 2015.
- [43] Y. Iraqi and A. Al-Dweik, "Power allocation for reliable SIC detection of rectangular QAM-based NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8355–8360, Aug. 2021.
- [44] F. Kara and H. Kaya, "BER performances of downlink and uplink NOMA in the presence of SIC errors over fading channels," *IET Commun.*, vol. 12, no. 15, pp. 1834–1844, Sep. 2018.
- [45] M. Vaezi and H. V. Poor, "NOMA: An information-theoretic perspective," in *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Cham, Switzerland: Springer, 2019, pp. 167–193.
- [46] M. Vaezi, R. Schober, Z. Ding, and H. V. Poor, "Non-orthogonal multiple access: Common myths and critical questions," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 174–180, Oct. 2019.
- [47] M. López-Benítez, "Performance analysis of SNR threshold-setting strategies for adaptive modulation and coding under fading channels," *Phys. Commun.*, vol. 30, pp. 154–166, Oct. 2018.
- [48] B. Verney. *Laminated Card 802.11n/HT and 802.11ac/VHT/MCS, SNR and RSSI*. Accessed: Sep. 2021. <https://www.wlanpros.com/resources/laminated-card-802-11nht-802-11acvht-mcs-snr-rssi/>
- [49] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.
- [50] J.-T. Liu and S. B. Gelfand, "Optimized decision-feedback equalization for convolutional coding with reduced delay," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1859–1866, Nov. 2005.
- [51] K. M. Ahmed, S. P. Majumder, and B. M. Rahman, "Performance analysis of an OFDM wireless communication system with convolutional coding," in *Proc. 11th Int. Conf. Adv. Commun. Technol.*, Feb. 2009, pp. 1078–1083.
- [52] S. Ahmed and M. Kawai, "Interleaving effects on BER fairness and PAPR in OFDMA system," *Telecommun. Syst.*, vol. 52, no. 1, pp. 183–193, Jan. 2013.
- [53] A. W. Nazar, S. A. Hassan, and H. Jung, "BER analysis of a NOMA enhanced backscatter communication system," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.



- [54] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.
- [55] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [56] *Technical Specification Group Radio Access Network User Equipment Radio Transmission and Reception. Part 1: Range 1 Standalone (Release 16)*, 3rd Generation Partnership Project, document 3GPP TS 38.101-12020.
- [57] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. 98, no. 3, pp. 403–414, 2015.
- [58] C. Clancy, J. Hecker, E. Stuntebeck, and T. O'Shea, "Applications of machine learning to cognitive radio networks," *IEEE Wireless Commun.*, vol. 14, no. 4, pp. 47–52, Aug. 2007.
- [59] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 6th ed. New York, NY, USA: Wiley, 1991.
- [60] B. Clerckx, Y. Mao, R. Schober, E. A. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, "Is NOMA efficient in multi-antenna networks? A critical look at next generation multiple access techniques," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1310–1343, 2021.
- [61] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. Int. Symp. Inf. Theory (ISIT)*, 2004, p. 174.
- [62] Y. Qi and M. Vaezi, "Secure transmission in MIMO-NOMA networks," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2696–2700, Dec. 2020.
- [63] A. D. Dabagh and D. J. Love, "Precoding for multiple antenna Gaussian broadcast channels with successive zero-forcing," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3837–3850, Jul. 2007.
- [64] M. Kobayashi and G. Caire, "An iterative water-filling algorithm for maximum weighted sum-rate of Gaussian MIMO-BC," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1640–1646, Aug. 2008.
- [65] A. Krishnamoorthy and R. Schober, "Uplink and downlink MIMO-NOMA with simultaneous triangularization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3381–3396, Jun. 2021.
- [66] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.
- [67] O. El Ayach, S. W. Peters, and R. W. Heath, Jr., "The feasibility of interference alignment over measured MIMO-OFDM channels," *IEEE Trans. Veh. Technol.*, vol. 59, no. 9, pp. 4309–4321, Nov. 2010.
- [68] R. Bahl, N. Gultai, K. R. Dandekar, and D. Jaggard, "Impact of pattern reconfigurable antennas on interference alignment over measured channels," in *Proc. IEEE Globecom Workshops*, Dec. 2012, pp. 557–562.
- [69] S. Begashaw, D. H. Nguyen, and K. R. Dandekar, "Enhancing blind interference alignment with reinforcement learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.
- [70] N. N. Moghadam, H. Farhadi, P. Zetterberg, M. N. Khormuji, and M. Skoglund, "Interference alignment: Practical challenges and test-bed implementation," in *Contemporary Issues in Wireless Communications*. London, U.K.: InTech, 2014, pp. 51–71.
- [71] C. M. Yetis, J. Fanjul, J. A. Garcia-Naya, N. N. Moghadam, and H. Farhadi, "Interference alignment testbeds," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 120–126, Oct. 2017.
- [72] J. W. Massey, J. Starr, S. Lee, D. Lee, A. Gerstlauer, and R. W. Heath, "Implementation of a real-time wireless interference alignment network," in *Proc. Conf. Rec. 46th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2012, pp. 104–108.
- [73] 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.
- [74] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, "Coordinated beamforming for multi-cell MIMO-NOMA," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 84–87, Jan. 2017.
- [75] S. A. Jafar, *Interference Alignment: A New Look at Signal Dimensions in a Communication Network*. New York, NY, USA: Now, 2011.
- [76] E. Hossain and Y. Al-Eryani, "Large-scale NOMA: Promises for massive machine-type communication," 2019, *arXiv:1901.07106*. [Online]. Available: <http://arxiv.org/abs/1901.07106>
- [77] N. Zhang, K. Cheng, and G. Kang, "A machine-learning-based blind detection on interference modulation order in NOMA systems," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2463–2466, Dec. 2018.
- [78] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.
- [79] K. P. K. Reddy, Y. Yeleswarapu, and S. J. Darak, "Performance evaluation of cumulant feature based automatic modulation classifier on USRP testbed," in *Proc. 9th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2017, pp. 393–394.
- [80] H. Guo, J. Xu, S. Zhu, and S. Wu, "Realtime software defined self-interference cancellation based on machine learning for in-band full duplex wireless communications," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Mar. 2018, pp. 779–783.
- [81] X. Zhang and M. Vaezi, "Multi-objective DNN-based precoder for MIMO communications," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4476–4488, Jul. 2021.



**YUE QI** (Graduate Student Member, IEEE) received the B.Eng. and M.Eng. degrees in electronic engineering from Xidian University, Xi'an, China, in 2015 and 2018, respectively. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Villanova University, Villanova, PA, USA. Her research interests include NOMA, MIMO systems, the IoT heterogeneous networks, green communication, and signal processing. She has received the IEEE Communication Society Student Grant at ICC Conference, in 2020.



**XINLIANG ZHANG** (Graduate Student Member, IEEE) received the B.Eng. and M.Eng. degrees in electrical engineering from Xidian University, Xi'an, China, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Villanova University, Villanova, PA, USA. His research interests include physical layer security, machine learning for wireless communications, MIMO networks, and signal processing.



**MOJTABA VAEZI** (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Amirkabir University of Technology (Tehran Polytechnic) and the Ph.D. degree in electrical engineering from McGill University.

He was a Researcher with Ericsson Research, Montreal, Canada. From 2015 to 2018, he was with Princeton University as a Postdoctoral Research Fellow and an Associate Research Scholar. He is currently an Assistant Professor of Electronics and

Communications Engineering (ECE) with Villanova University. His research interests include the broad areas of signal processing and machine learning for wireless communications with an emphasis on physical layer security and fifth-generation (5G), and beyond radio access technologies. Among his publications in these areas is the book *Multiple Access Techniques for 5G Wireless Networks and Beyond* (Springer, 2019). He was a recipient of several academic, leadership, and research awards, including McGill Engineering Doctoral Award, the IEEE Larry K. Wilson Regional Student Activities Award, in 2013, the Natural Sciences and Engineering Research Council of Canada (NSERC) Postdoctoral Fellowship, in 2014, the Ministry of Science and ICT of Korea's Best Paper Award, in 2017, the IEEE Communications Letters Exemplary Editor Award in 2018, the 2020 IEEE Communications Society Fred W. Ellersick Prize, and the 2021 IEEE Philadelphia Section Delaware Valley Engineer of the Year Award. He has co-organized six NOMA workshops at IEEE VTC 2017-Spring, Globecom 2017 and Globecom 2018, and ICC 2018–2020. He is an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE COMMUNICATIONS LETTERS.

...