Beamforming Design and Power Allocation for Secure Transmission With NOMA
Youhong Feng, Member, IEEE, Shihao Yan, Member, IEEE, Zhen Yang, Nan Yang, Senior Member, IEEE, and Jinhong Yuan, Fellow, IEEE

Abstract—In this paper, we propose a novel beamforming design to enhance physical layer security of a non-orthogonal multiple access (NOMA) system with the aid of artificial noise (AN). The proposed design uses two factors to balance the useful signal strength and interference at the strong and weak users, which is a generalized version of the existing beamforming designs in the context of physical layer security for NOMA. We determine the optimal power allocation among useful signals and AN together with the two optimal factors in order to maximize the secrecy sum rate (SSR). Our asymptotic analysis in the high signal-to-noise ratio regime provides an efficient and near-optimal solution to optimize the beamforming scalars and power allocation coefficients. Our analysis indicates that it is not optimal to form a beam toward either the strong user or the weak user in NOMA systems for security enhancement. In addition, the asymptotically optimal power allocation informs that, as the transmit power increases, more power should be allocated to the weak user or AN signals, while the power allocated to the strong user keeps constant. Our examination shows that the proposed novel beamforming design can significantly outperform two benchmark schemes.

Index Terms—Non-orthogonal multiple access, physical layer security, artificial noise, optimal power allocation.

I. INTRODUCTION

NON-ORTHO GONAL multiple access (NOMA), as a potentially promising technique to significantly boost the system spectral efficiency in the fifth-generation (5G) and beyond wireless networks, has attracted an increasing amount of research effort [1]–[7]. Different from the conventional orthogonal multiple access (OMA) techniques, such as frequency division multiple access, time division orthogonal multiple access, and code division multiple access, NOMA can exploit the power domain to serve multiple users simultaneously in the same resource block (i.e., time/frequency/code). Motivated by the improved spectral efficiency provided by NOMA, different issues in NOMA systems have been addressed in the literature (e.g., [5], [8]–[13]). For example, [8] focused on a downlink NOMA system, where the authors considered user pairing and transmit power allocation to enhance the performance of NOMA. In [9], an optimal transmit power allocation scheme was proposed in multiple-input multiple-output (MIMO) NOMA systems in order to maximize the sum rate of two paired users subject to some specific constraints. In addition, the authors of [5] proposed a joint subcarrier and power allocation scheme to maximize the weighted sum rate of a NOMA system. Considering delay constraint, the authors of [10] tackled the maximization of the effective throughput in the context of NOMA for short-data communications, which shows that NOMA can aid to achieve low-latency communications. Furthermore, in [4], the authors have surveyed the recent literature of power-domain multiplexing-aided NOMA proposed for 5G systems and then provided an extensive work on NOMA. Particularly, interference management as an important technology has a vital impact on the performance of NOMA [4]. For example, successive interference cancellation (SIC) is a key enabler of the power-domain NOMA since it effectively suppresses the interference with low complexity [14]–[16]. The use of SIC has been extensively investigated in wireless networks to provide all the users with acceptable quality of services [4], [17].

Wireless communication security is another issue of growing importance in 5G and beyond wireless networks, since there is an increasing amount of confidential information
(e.g., credit card information) that is transferred over the air. Physical layer security, as a complementary and alternative cryptographic method to defend against eavesdroppers, exploits the inherent properties (e.g., randomness) of the wireless medium to achieve the ever-lasting and information-theoretic secrecy (e.g., [18]–[25]). In this context, MIMO architectures (e.g., [20], [21]) and artificial-noise (AN)-aided secure transmissions (e.g., [21]–[25]) have been widely adopted to enhance the secrecy performance of wireless communications. Against this background, physical layer security in NOMA systems has been partially addressed [26]–[35]. For example, in [28] the authors considered physical layer security in a single-input single-output (SISO) NOMA system and proposed an optimal power allocation policy for maximizing the secrecy sum rate (SSR) of all users subject to their predefined quality of service (QoS) requirements. In [29], the authors focused on the transmission power minimization in a multiple-input single-output (MISO) NOMA cognitive radio network in the presence of multiple single-antenna eavesdroppers. Considering a cell-edge user (i.e., the weak user) as a potential eavesdropper to an entrusted central user (i.e., the strong user), the maximization of the secrecy rate of the central user subject to a transmit power constraint and a transmission rate requirement at the cell-edge user is tackled in [30]. In [31], the authors focused on the SSR optimization problem for a downlink MIMO NOMA network subject to successful SIC and transmit power constraints, in which the nonconvex maximization of the SSR was transformed to a biconvex problem and solved by the alternating optimization method.

In the NOMA systems considered in [29]–[31], either the perfect knowledge on the eavesdropper’s instantaneous channel state information (CSI) or a bounded error model on the eavesdropper’s instantaneous CSI was considered. Such CSI information may not be achievable in some specific application scenarios of NOMA, in which the eavesdropper is not an internal user or an active receiver. As such, the assumption that only the statistical information on the eavesdropper’s CSI (e.g., a passive eavesdropping scenario) was widely used in the context of physical layer security for NOMA (e.g., [32]–[34]). Specifically, [32] proposed a NOMA scheme that maximizes the minimum confidential information rate under the secrecy outage probability (SOP) and transmit power constraints. Inspired by the enhanced secrecy performance achieved by AN-aided transmission strategies, [33] and [34] considered AN-aided secure beamforming (SBF) strategies to protect the confidential information of legitimate users for MISO NOMA systems. More specifically, the authors of [33] considered large-scale networks with randomly deployed legitimate users and eavesdroppers, where the exact and asymptotic expressions for the SOP were derived. The imperfect SIC was considered in [34], where the SOPs of the legitimate users were obtained in closed-form expressions.

As for the SBF design in NOMA systems, in [33] a maximum ratio transmission (MRT) strategy was adopted, i.e., the weak user (User 1) and the strong user (User 2) adopted two different beamforming vectors to transmit useful signals to the weak user and the strong user. In this MRT strategy, the signal strengths of $s_1$ and $s_2$ are maximized at the weak user and the strong user, respectively. As clarified in [34], this MRT strategy may not guarantee perfect SIC at User 2, since the interference caused by $s_2$ is also maximized when User 2 decodes $s_1$ to conduct SIC. Thus, in [34] the same SBF (i.e., the two beamforming vectors to transmit useful signals to the weak user and the strong user are the same) was designed such that the signal strengths of both $s_1$ and $s_2$ are maximized at User 2. In the SBF designs proposed by [33] and [34], we observe that the balance between the useful signal strength and interference was not struck, i.e., the useful signal strengths and interference are either minimized or maximized. We note that this balance can potentially enhance the achieved physical layer security in NOMA systems, since the useful signal $s_1$ should be decoded at both User 1 and User 2, while the useful signal $s_2$ causes interference at both User 1 and User 2 (but $s_2$ is only decoded at User 2). Thus, the motivation of this work is to design a new SBF scheme which achieves the balance between the useful signal strength and interference such that the secrecy performance of NOMA systems is improved.

- We propose a novel hybrid SBF scheme in a NOMA system to enhance physical layer security by balancing the useful signal strength and interference at both User 1 and User 2. Specifically, in our proposed scheme the beamforming vectors for the weak user and strong user are linear functions of the channel vector of the weak user, the channel vector of the strong user, and a random vector. In this scheme, AN is also used to further enhance the secrecy performance of NOMA systems. Thus, we refer to this scheme as the NOMA-HB-AN scheme. In this scheme, the beams used to transmit $s_1$ and $s_2$ can be in any direction (e.g., possibly towards neither User 1 nor User 2). We note that the proposed NOMA-HB-AN scheme is a generalized version of the SBF schemes proposed in [34], [35], [38], and [39].

- In order to maximize the benefits of the proposed NOMA-HB-AN scheme, we tackle the optimization of two governing parameters $\beta_1$ and $\beta_2$ (i.e., $\beta_1$ and $\beta_2$ are beamforming parameters to control the weak user’s and the strong user’s beamforming vectors, respectively) together with the optimal power allocation among $s_1$, $s_2$, and the AN signals, aiming to maximize the SSR subject to specific QoS constraints at the two legitimate users. Considering a larger number of transmit antennas, we first determine the optimal power allocation for given $\beta_1$ and $\beta_2$, in which the power allocation coefficients for $s_1$ and $s_2$ are analytically derived as functions of the power allocation coefficient for AN signals. This leads to the optimal power allocation can be achieved with the aid of a one-dimensional numerical search. Our results show that the proposed NOMA-HB-AN scheme can significantly outperform the SBF design with the same beamforming vector proposed in [34].

- To gain further insights into the proposed scheme, we consider the joint optimization of $\beta_1$ and $\beta_2$ together with power allocation in the high signal-to-noise (SNR) regime. Particularly, we derive the power allocation coeffi-
Fig. 1. Illustration of a downlink MISOME NOMA system, where the BS is equipped with $N$ antennas, each of User 1 and User 2 is equipped with a single antenna, and the eavesdropper is equipped with $K$ antennas.

coefficients for $s_1$, $s_2$, and the AN signals in closed-form expressions at high SNRs, based on which the optimization of $\beta_1$ and $\beta_2$ can be efficiently achieved by another one-dimensional numerical search. Our results show that the achieved optimal $\beta_1$, $\beta_2$, and power allocation in the high SNR regime can precisely approximate those achieved for arbitrary SNRs, in terms of achieving similar maximum SSRs. This indicates that our proposed NOMA-HB-AN scheme can be efficiently optimized and the incurred complexity increase is negligible. Our results also show that, when the eavesdropper’s channel quality is not high, the careful design of SBF is more important than simply using AN into the design, which is confirmed by our observation that the proposed NOMA-HB-AN scheme without AN can even outperform the SBF design with the same beamforming vector and AN [34].

The remainder of this paper is organized as follows. In Section II, the system model and the hybrid SBF are presented. The maximization of the SSR under the QoS constraints at the two legitimate users is formulated in Section III. The solution to the SSR maximization problem are provided in Section IV, where the scenarios with arbitrary and high SNRs are considered. Numerical results are provided in Section V to offer valuable insights on the secrecy performance of the proposed scheme compared with two benchmark schemes. Conclusions are drawn in Section VI.

Notation: Scalar variables are denoted by italic symbols; Vectors and matrices are denoted by lower-case and upper-case boldface symbols, respectively; $A^H$ denotes the Hermitian (conjugate) transpose of a matrix $A$; $I_K$ represents the $K \times K$ identity matrix; $E[x]$ denote the mean of the random variable $x$; $x \sim CN(\mu, \sigma^2)$ denotes a circularly symmetric complex Gaussian random variable $x$ with mean $\mu$ and covariance $\sigma^2$.

II. SYSTEM MODEL

As shown in Fig. 1, we consider the secure transmission using NOMA from a base station (BS) to two legitimate users in the presence of a multi-antenna eavesdropper (Eve). The BS is equipped with $N$ antennas, each of the legitimate users (i.e., User 1 and User 2) is equipped with a single antenna, and Eve is equipped with $K$ antennas. As such, we refer to the considered system as a MISOME NOMA system. We assume that $N$ is large, and in particular is much larger than $K$. The channel vector from the BS to the legitimate user $m \in \{1, 2\}$ is denoted by $h_m \in \mathbb{C}^{1 \times N}$, of which the entries are independent and identically distributed (i.i.d.) circularly-symmetric complex Gaussian random variables with zero-mean and variance $\sigma_m^2$. The channel matrix from the BS to Eve is denoted by $H_e \in \mathbb{C}^{K \times N}$, where $h_{e,k} \triangleq H_e(k,:)$ and $h_{e,k} \in \mathbb{C}^{1 \times N}$ is an $1 \times N$ channel vector from the BS to the $k$-th receive antenna at Eve and $e_k$ is a random vector following a complex circular Gaussian distribution with mean $0$ and covariance $\delta_e^2$.

In this work, we assume that the CSI of all the legitimate channels (i.e., $h_m$) is known at the BS, while only the statistical CSI of the Eve’s channel (i.e., the statistical information on $H_e$) is available. It is a very generic assumption that the statistical CSI of the Eve’s channel is known, which has been widely adopted in the literature of physical layer security [34], [39], [40]. Without loss of generality, we assume that the legitimate channel gains are sorted in ascending order [32], [35], i.e., $0 < \|h_1\|^2 \leq \|h_2\|^2$.

A. Secure Transmission With NOMA and Artificial Noise

We next detail the secure transmission using NOMA with AN in our considered system. Specifically, the BS transmits two information signals, $s_1$ and $s_2$, in conjunction with an $(N-2) \times 1$ AN vector $s_N$ to its corresponding receivers, where $s_m$ is the information signal dedicated for the $m$-th user. The variance of $s_m$ is denoted by $\chi_m$ and the total transmit power is denoted by $P$. We denote $\phi_m$ as the power allocation coefficient to $s_m$, where $0 < \phi_m \leq 1$, which determines the fraction of the total transmit power allocated to $s_m$ such that $\chi_m = \phi_m P$. Since the BS does not know $H_e$, it equally distributes the AN transmit power to each entry of $s_N$ and thus the variance of each entry of $s_N$ is the same, which is denoted by $\chi_N$. Then the BS transmits $s_N$ in the null space of the channel from the BS to the two users, $H \triangleq [h_1^H, h_2^H]$, such that $s_N$ leads to interference at Eve but not at the two legitimate users. As such, we know that all the remaining transmit power (excluding the power allocated to $s_1$ and $s_2$) should be used to transmit $s_N$, such that we have $\chi_N = \phi_e P/(N-2)$ with $\phi_e = 1 - \phi_1 - \phi_2$. To transmit $s_m$ and $s_N$, the BS has to design an $N \times N$ beamforming matrix $V$ given by

$$V = [v_1, v_2, V_N],$$

where we recall that $v_1$ and $v_2$ are the beamforming vectors used to transmit $s_1$ and $s_2$, respectively, and $V_N$ is the unitary beamforming matrix used to transmit $s_N$. In this work, we adopt specific structures for $v_1$ and $v_2$ given as

$$v_1 = \frac{\sqrt{\beta_1}h_1 + \sqrt{(1-\beta_1)}h_2}{\|\sqrt{\beta_1}h_1 + \sqrt{(1-\beta_1)}h_2\|},$$

$$v_2 = \frac{\sqrt{\beta_2}h_2 + \sqrt{(1-\beta_2)}e}{\|\sqrt{\beta_2}h_2 + \sqrt{(1-\beta_2)}e\|},$$

where $\beta_1$ and $\beta_2$ are design parameters to be determined later, $h_i = h_i^H/\|h_i\|$, $e = \mathcal{CN}(0, \delta_e^2)$, and $e \in \mathbb{C}^{1 \times N}$ is a random vector that does not align with $h_1$ or $h_2$. The design of $v_1$
originates from the fact that the information signal $s_1$ need to be decoded by both User 1 and User 2 (User 2 decodes $s_1$ by performing SIC) and the design of $v_2$ originates from that only User 2 decodes $s_2$ while $s_2$ causes interference at User 1 for decoding $s_1$. We note that the proposed $v_1$ and $v_2$ are generalizations of the beamforming vectors adopted in existing works (e.g., [33], [34]) and thus they are expected to achieve better system performance with optimized $\beta_1$ and $\beta_2$, which will be confirmed by our examination in this work.

Using $V$, the transmitted signal vector at the BS is given by

$$s = V \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = v_1 s_1 + v_2 s_2 + V_N s_N. \quad (4)$$

Therefore, the received signal at the $m$-th user is given by

$$y_m = h_m s + n_m = h_m \sum_{i=1}^{2} v_i s_i + h_m V_N s_N + n_m = h_m \sum_{i=1}^{2} v_i s_i + n_m, \quad (5)$$

where $n_m$ satisfying $E[n_m n_m^H] = \sigma_n^2 I_m$ is the additive white Gaussian noise (AWGN) at the $m$-th user and $h_m V_N = 0$ is applied, since $s_N$ is transmitted in the null space of $H$. Likewise, the received signal vector at Eve is given by

$$y_e = H_e s + n_e = H_e \sum_{i=1}^{2} v_i s_i + H_e V_N s_N + n_e, \quad (6)$$

where $n_e$ satisfying $E[n_e n_e^H] = \sigma_e^2 I_K$ is the AWGN vector at Eve.

B. Performance Metric for NOMA With AN

According to the principle of NOMA, the user with the better channel condition (i.e., User 2) first decodes the signal of the other user (i.e., User 1) and then successively subtracts the interference caused by this signal from its received signal before decoding its own information. The weak user, User 1 directly decodes its own information by treating User 2’s signal as interference [41]. As such, the maximum achievable rate of $s_1$ is given by [34] and [36]

$$R_{u1} = \log_2 (1 + \gamma_{u1}), \quad (7)$$

where $\gamma_{u1}$ denotes the achievable SINR of $s_1$ with

$$\gamma_{u1} = \min \left( \frac{\phi_1 P |h_1 v_1|^2}{\phi_2 P |h_2 v_1|^2 + \sigma_1^2}, \frac{\phi_1 P |h_2 v_2|^2}{\phi_2 P |h_2 v_2|^2 + \sigma_2^2} \right), \quad (8)$$

while the first and second terms on the right-hand side of (8) denote the received SINR for decoding User 1’s signal $s_1$ at User 1 and User 2, respectively. We note that the “min” function used in (8) comes from the assumption that perfect SIC is guaranteed at User 2, which is assumed in this work. With SIC, User 2 decodes its information signal without interference and thus, the maximum achievable rate of $s_2$ is given by

$$R_{u2} = \log_2 (1 + \gamma_{u2}), \quad (9)$$

where $\gamma_{u2} = \phi_2 P |h_2 v_2|^2 / \sigma_2^2$ denotes the SNR for decoding User 2’s signal at User 2.

In this work, we consider the worst-case scenario, where Eve has already decoded the information signal for User 1 in order to conduct SIC before it attempts to decode the information for User 2, which is exactly the same as the decoding procedure at User 2 (i.e., the strong user). The worst-case assumption has been widely adopted in designing and analyzing the NOMA transmission schemes with physical layer security (e.g., [28], [32]). As such, the maximum achievable rate of $s_1$ at Eve is given by [42], [43]

$$R_{e1} = \log_2 \det \left( \sigma_e^2 I_K \right) + \frac{\phi_1 P H_e v_1 (H_e v_1)^H}{\phi_2 P H_e v_2 (H_e v_2)^H + \frac{\sigma_e^2}{\sigma_n^2} H_e V_N (H_e V_N)^H + \sigma_e^2 I_K} \right). \quad (10)$$

Likewise, the maximum achievable rate of $s_2$ at Eve is given by [42], [43]

$$R_{e2} = \log_2 \det \left( \sigma_e^2 I_K + \frac{\phi_2 P H_e v_2 (H_e v_2)^H}{\frac{\sigma_e^2}{\sigma_n^2} H_e V_N (H_e V_N)^H + \sigma_e^2 I_K} \right). \quad (11)$$

We denote the achievable secrecy rate of $s_m$ and the SSR as $R_{sm}$ and $R_s$, respectively. Therefore, we have

$$R_s = \sum_{m=1}^{2} R_{sm} = \sum_{m=1}^{2} \left[ R_{um} - R_{em} \right]^+, \quad (12)$$

where $[x]^+ = \max(0, x)$. We note that our initial motivation for considering a sufficiently large number of transmit antennas in this work is to achieve tractable analysis in order to provide guidelines on the development of secrecy beamforming strategy in the context of NOMA. Specifically, as the number of transmit antennas becomes sufficiently large, the eavesdropper’s channel converges to its mean value and thus the secrecy outage occurs with a negligible probability, which allows us to adopt the SSR as a performance metric in the considered NOMA system and offers some tractable analysis [44], [45].

Although the authors of [33], [34], [37], [38] analyzed the secrecy performance of the NOMA system with multiple users and eavesdroppers, they considered simple maximum ratio transmission (MRT) beamforming (i.e., $v_i = h_i / |h_i|$ in [37], [38], and [33] and $v_i = h_2 / |h_2|$ in [34]). This consideration leads to the loss of the freedom to vary the tradeoff between the useful signal strength and interference strength, i.e., the useful signal strength and interference are either minimized or maximized. In addition, only fixed power allocation was considered in [38] and no artificial noise-aided scheme was employed in [33]. This also limits the possibility of improving the secrecy performance of the considered NOMA system. In order to achieve the freedom to balance the useful signal strength and interference strength, we develop a new
AN-aided SBF (secrecy beamforming) scheme and optimize its associated controlling parameters to further improve the secrecy performance of the NOMA systems.

We note that the two-user NOMA system is the most typical NOMA scenario in the literature (e.g., [9], [26], [33], [34], [36], [46]–[49]), which is motivated by some practical applications such as multi-user superposition transmission (MUST) in the third-generation partnership project long-term evolution (3GPP-LTE) [34]. We note that the developed beamforming strategy in this work can be extended into a scenario with more than two users, where user pairing can be applied to construct a hybrid NOMA system (e.g., [34], [46], [50]). Specifically, a group of users can be divided into multiple subgroups of users according to some certain criteria, e.g., based on the channel conditions, the spatial correlation of channels, or the power differences between users. This allows us to apply the developed beamforming strategy to suppress or minimize the interference between subgroups in order to achieve the benefits of NOMA within each subgroup of users. This issue will be studied as a research topic in our future works.

III. OPTIMIZATION FRAMEWORK WITH A SUFFICIENT LARGE NUMBER OF TRANSMIT ANTENNAS

In this section, we first present the adopted optimization framework. By considering a sufficiently large number of transmit antennas, we conduct new analysis to simplify the objective function and the corresponding constraints.

A. Optimization Framework

In this work, following [28], [35] we aim to maximize the SSR (i.e., $R_s$) subject to some constraints on the maximum achievable rates of $s_1$ and $s_2$ (i.e., $R_{u1}$ and $R_{u2}$). Specifically, the focused optimization problem can be written as

$$\textbf{P1} : \max_{\phi_1, \phi_2, \phi_e, \beta_1, \beta_2} R_s$$

s.t. $R_{um} \geq Q_m, ~ m \in \{1, 2\}$, 

$$\phi_1 + \phi_2 + \phi_e = 1,$$

$$0 \leq \beta_m \leq 1, ~ m \in \{1, 2\},$$

where $Q_m$ denotes the minimum codeword rate required by the $m$-th legitimate user. We note that the constraint given in (14) can be justified by the fact that secure transmission is only considered when the QoS without security at the $m$-th user is above a specific threshold [28], [35], [51]. We also note that in the optimization problem given in (13) we have $0 < \phi_m < 1$ and $0 \leq \phi_e < 1$.

Due to the constraint given in (14), there exists a minimum transmit power, denoted by $P_{\min}$, that guarantees the feasibility of the optimization problem $\textbf{P1}$. In other words, the optimization problem $\textbf{P1}$ is feasible only when $P \geq P_{\min}$. The value of $P_{\min}$ can be determined following the method given in [28] and thus in this work we assume that this feasible condition is always guaranteed.

B. SINR of $s_1$ and Constraint $R_{um} \geq Q_m$

In this subsection, we present the determined expression (without “min”) for the SINR of $s_1$ in the following lemma to facilitate solving the optimization problem $\textbf{P1}$, based on which we also transfer the constraint $R_{um} \geq Q_m$ into a specific constraint on $\phi_m$.

Lemma 1: In the solution to the optimization problem $\textbf{P1}$, the achieved SINR for $s_1$ is given by

$$\gamma_{u1} = \frac{\phi_1 P|h_1 v_1|^2}{\phi_2 P|h_1 v_2|^2 + \sigma_1^2} = \frac{\phi_1 P|h_2 v_1|^2}{\phi_2 P|h_2 v_2|^2 + \sigma_2^2}. \quad (17)$$

Proof: Following (8), in order to prove this lemma we only have to prove that

$$\frac{\phi_1 P|h_1 v_1|^2}{\phi_2 P|h_1 v_2|^2 + \sigma_1^2} = \frac{\phi_1 P|h_2 v_1|^2}{\phi_2 P|h_2 v_2|^2 + \sigma_2^2} \quad (18)$$

is always guaranteed in the solution to the optimization problem $\textbf{P1}$. In what follows, we prove (18) by contradiction. We first assume that

$$\frac{\phi_1 P|h_1 v_1|^2}{\phi_2 P|h_1 v_2|^2 + \sigma_1^2} > \frac{\phi_1 P|h_2 v_1|^2}{\phi_2 P|h_2 v_2|^2 + \sigma_2^2}. \quad (19)$$

holds in the solution to the optimization problem $\textbf{P1}$. As per (8), following (19) we have

$$\gamma_{u1} = \frac{\phi_1 P|h_2 v_1|^2}{\phi_2 P|h_2 v_2|^2 + \sigma_2^2}. \quad (20)$$

Based on (2), in this case we can decrease $\beta_1$ in order to increase $\gamma_{u1}$ by slightly increasing the right-hand-side of (19) while decreasing its left-hand-side. This leads to the increase in the SSR (i.e., $R_s$), which is given in (12), which is contradict to the assumption that (19) is guaranteed in the solution to the optimization problem $\textbf{P1}$. We have a similar argument for the assumption of

$$\frac{\phi_1 P|h_1 v_1|^2}{\phi_2 P|h_1 v_2|^2 + \sigma_1^2} < \frac{\phi_1 P|h_2 v_1|^2}{\phi_2 P|h_2 v_2|^2 + \sigma_2^2}, \quad (21)$$

where we can increase the SSR $R_s$ by increasing $\beta_1$. As such, we complete the proof of Lemma 1.

Following Lemma 1, for clarity in this work we write the SINR of $s_1$ as

$$\gamma_{u1} = \frac{\phi_1 P|h_1 v_1|^2}{\phi_2 P|h_1 v_2|^2 + \sigma_1^2}. \quad (22)$$

Following (7), (9), and Lemma 1, for given $\beta_1$ and $\beta_2$, the constraint $R_{um} \geq Q_m$ given in (14) can be rewritten as

$$\phi_1 \geq \frac{2^{Q_1} - 1}{P|h_1 v_1|^2} \left(\phi_2 P|h_1 v_2|^2 + \sigma_1^2\right), \quad (23)$$

and

$$\phi_2 \geq \frac{2^{Q_2} - 1}{P|h_2 v_2|^2} \sigma_2^2, \quad (24)$$

respectively.
C. Secrecy Sum Rate With Sufficiently Large $N$

Considering $N \to \infty$, we present an approximated but closed-form expression for the SSR (i.e., the objective function in the optimization problem $P_1$) in the following theorem.

**Proposition 1:** As $N \to \infty$ with $N \gg K$, the SSR given in (12) can be approximated as

$$
\tilde{R}_s = \log_2 \left( 1 + \frac{(N-1)\beta_1 + 1 + \phi_1 \rho_{u_1}}{\phi_2 \rho_{u_1} + 1} \right) \\
\times \left( 1 + ((N-1)\beta_2 + 1 + \phi_2 \rho_{u_2}) \right) + \log_2 \left( 1 + (1 - \phi_1 - \phi_2) \rho_e \right)^K - K \log_2(1 + \rho_e)
$$

where $\rho_{u_1} = P \delta_1^2 / \sigma_1^2$, $\rho_{u_2} = P \delta_2^2 / \sigma_2^2$, and $\rho_e = P \delta_e^2 / \sigma_e^2$ are the average SNRs of the BS-User 1, BS-User 2, and BS-Eve links, respectively.

**Proof:** The proof is presented in Appendix A.

In the remaining of this work, we use the approximated SSR (i.e., $\tilde{R}_s$) instead of $R_s$ as our objective function, since we focus on the scenario with a sufficiently large number of transmit antennas.

**Lemma 2:** As $N \to \infty$, the constraints in (23) and (24) can be rewritten as

$$
\phi_1 \geq \frac{2^{Q_1} - 1}{((N-1)\beta_1 + 1) \rho_{u_1}} (1 + \phi_2 \rho_{u_1}),
$$

and

$$
\phi_2 \geq \frac{2^{Q_2} - 1}{((N-1)\beta_2 + 1) \rho_{u_2}},
$$

respectively.

**Proof:** The proof of Lemma 2 follows similar arguments as that of Proposition 1 and thus is omitted here.

Based on Proposition 1 and Lemma 2, we focus on the optimization problem $P_2$, instead of $P_1$, in the remaining of this work. Specifically, $P_2$ is expressed as

$$
P_2: \max_{\phi_1, \phi_2, \phi_e, \beta_1, \beta_2} \tilde{R}_s
$$

s.t.

$$
\phi_1 \geq \frac{(2^{Q_1} - 1)(1 + \phi_2 \rho_{u_1})}{((N-1)\beta_1 + 1) \rho_{u_1}},
$$

$$
\phi_2 \geq \frac{2^{Q_2} - 1}{((N-1)\beta_2 + 1) \rho_{u_2}},
$$

$$
\phi_1 + \phi_2 + \phi_e = 1,
$$

$$
0 \leq \beta_m \leq 1, \quad m \in \{1, 2\}.
$$

We will tackle the optimization problem $P_2$ in the following section.

IV. POWER ALLOCATION AND BEAMFORMING DESIGN IN MISOME NOMA SYSTEMS

In this section, we first solve $P_2$ for given values of $\beta_1$ and $\beta_2$, where we recast the SSR maximization as a two-level optimization framework that involves a one-dimensional numerical search. Then, we analytically determine the optimal power allocation in the high SNR regime for fixed $\beta_1$ and $\beta_2$. Finally, we provide the method for obtaining the optimal $\beta_1$ and $\beta_2$.

A. Optimal Power Allocation for Given $\beta_1$ and $\beta_2$

For given $\beta_1$ and $\beta_2$, the optimization problem $P_2$ can be rewritten as

$$
P_3 : \max_{\phi_1, \phi_2, \phi_e} \tilde{R}_s(\beta_1, \beta_2) = \max_{\phi_e} \left\{ \log_2 (1 + \phi_e \rho_{e})^K \right\}
$$

$$
+ \max_{\phi_1, \phi_2, \phi_e} \log_2 \left[ \left( 1 + \frac{\phi_1 \rho_{u_1}}{1 + \phi_2 \rho_{u_1}} \right) \left( 1 + \phi_2 \rho_{u_2} \right) \right] - K \log_2(1 + \rho_e),
$$

s.t.

$$
\phi_1 = 1 - \phi_1 - \phi_e,
$$

$$
\phi_1 \geq \frac{1}{c_1} (2^{Q_1} - 1)(1 + \phi_2 \rho_{u_1}),
$$

$$
\phi_2 \geq \frac{1}{c_2} (2^{Q_2} - 1),
$$

where $\tilde{R}_s(\beta_1, \beta_2)$ denotes $\tilde{R}_s$ given in (28) for given $\beta_1$ and $\beta_2$, $c_1 = 1 + (N-1)\beta_1 \rho_{u_1}$, and $c_2 = 1 + (N-1)\beta_2 \rho_{u_2}$.

Due to the high complexity of the objective function in the optimization problem $P_3$, we solve it in the following two steps. In the first step, for a given power allocation coefficient $\phi_e$, we obtain closed-form expressions for the optimal values of the power allocation coefficients $\phi_1$ and $\phi_2$, which are functions of $\phi_e$. In the second step, we adopt a one-dimensional numerical search to determine the optimal value of $\phi_e$, which leads to the optimal power allocation for given $\beta_1$ and $\beta_2$.

In the first step of solving the optimization problem $P_3$, we tackle the optimization problem for a given $\phi_e$, which is

$$
P_4 : \max_{\phi_1, \phi_2} F(\phi_1, \phi_2)
$$

s.t.

$$
\phi_1 + \phi_2 = 1 - \phi_e,
$$

$$
\phi_1 \geq \frac{1}{c_1} (2^{Q_1} - 1)(1 + \phi_2 \rho_{u_1}),
$$

$$
\phi_2 \geq \frac{1}{c_2} (2^{Q_2} - 1),
$$

where

$$
F(\phi_1, \phi_2) = \log_2 \left[ \left( 1 + \frac{\phi_1 \rho_{u_1}}{1 + \phi_2 \rho_{u_1}} \right) \left( 1 + \phi_2 \rho_{u_2} \right) \right].
$$

We note that the feasible range of $\phi_e$ is $0 \leq \phi_e \leq P_{\text{min}}$, where we recall that $P_{\text{min}}$ is the minimum transmit power that guarantees the QoS constraints at the two legitimate users. We also note that due to the constraint given in (38) the only parameter to optimize in $P_4$ is $\phi_1$ or $\phi_2$. Here, we take $\phi_2$ as the parameter to optimize. Then, we have the following lemma to facilitate solving the optimization problem $P_4$.

**Lemma 3:** For $\phi_1 + \phi_2 = 1 - \phi_e$, the objective function in $P_4$, i.e., $F(\phi_1, \phi_2)$ given in (41), is a concave function of $\phi_2$.

**Proof:** The proof is presented in Appendix B.

Following Lemma 3, the solution to the optimization problem $P_4$ is given in the following theorem.

**Theorem 1:** For a given feasible $\phi_e$, the optimal values of $\phi_1$ and $\phi_2$ for the optimization problem $P_4$ are derived as
functions of $\phi_e$, given by

$$\phi_2^+(\phi_e) = \begin{cases} 
\mu_0, & \text{when } \mu_1 \leq \mu_0 \leq \mu_2, \\
\mu_1, & \text{when } \mu_0 < \mu_1, \\
\mu_2, & \text{when } \mu_0 > \mu_2,
\end{cases} \quad (42)$$

$$\phi_2^-(\phi_e) = 1 - \phi_e - \phi_2^+(\phi_e), \quad (43)$$

where

$$\mu_0 = \frac{\sqrt{c_2^2 c_3^2 - c_2 c_3 p_u c_4 - c_2 c_3}}{c_2 c_3 p_u}, \quad (44)$$

$$\mu_1 = \frac{1}{c_2} (2^{Q_2} - 1), \quad (45)$$

$$\mu_2 = \frac{1 - \phi_e - \frac{1}{c_1} (2^{Q_1} - 1)}{1 + \frac{(2^{Q_1} - 1)}{(N-1)\beta_1 + 1}}. \quad (46)$$

**Proof:** Based on Lemma 3, the optimal value of $\phi_2$ that maximizes the objective function in $P4$ without considering the constraints given in (39) and (40) is the one that guarantees $\frac{\partial F(\phi_1, \phi_2)}{\partial \phi_2} = 0$ (i.e., $G(\phi_2) = 0$ in (75) of Appendix B), which is given by (following Lemma 3 again)

$$\mu_0 = \frac{\sqrt{c_2^2 c_3^2 - c_2 c_3 p_u c_4 - c_2 c_3}}{c_2 c_3 p_u}. \quad (47)$$

Substituting (38) into (39), the constraints given in (39) and (40) can be rewritten as the constraints on $\phi_2$, given by

$$\mu_1 \leq \phi_2 \leq \mu_2. \quad (48)$$

If $\mu_0$ satisfies the constraints given in (48), we can directly conclude $\phi_2^+(\phi_e) = \mu_0$. Otherwise, we have the following two cases. For $\mu_0 < \mu_1$, we have $\phi_2^+(\phi_e) = \mu_1$. This is due to the fact that, as we proved in Lemma 3, the objective function $F(\phi_1, \phi_2)$ is a concave function of $\phi_2$ and $\mu_0$ is the value of $\phi_2$ that maximizes $F(\phi_1, \phi_2)$, which leads to the fact that when $\mu_0 < \mu_1$ the objective function $F(\phi_1, \phi_2)$ monotonically decreases with $\phi_2$ for $\mu_1 \leq \phi_2 \leq \mu_2$. Following a similar argument, we have $\phi_2^-(\phi_e) = \mu_2$ when $\mu_0 > \mu_2$. This completes the proof of Theorem 1.

Following Theorem 1, in the second step of solving the optimization problem $P3$, we have to solve a univariate optimization problem with respect to $\phi_e$, which is given by

$$P5 : \max_{0 \leq \phi_e \leq \frac{P - P_{min}}{P}} \left[ \log_2 (1 + \phi_e \rho_e)^K + F(\phi_e^*(\phi_e), \phi_e^*(\phi_e)) \right] - K \log_2 (1 + \rho_e). \quad (49)$$

We note that the optimization problem $P5$ is identical to $P3$. For the optimization problem $P5$, we can perform a one-dimensional numerical search over $0 \leq \phi_e \leq \frac{P - P_{min}}{P}$ to determine the optimal value of $\phi_e$, i.e., $\phi_e^*$. Then, substituting $\phi_e^*$ into Theorem 1 we can obtain the optimal values of $\phi_1$ and $\phi_2$, which are denoted by $\phi_1^*$ and $\phi_2^*$, respectively. So far, we have solved the optimization problem $P3$ with the aid of a one-dimensional numerical search, which determines the optimal power allocation strategy for given $\beta_1$ and $\beta_2$. In order to reduce the complexity of determining the optimal power allocation and provide some insights based on analysis, in the following subsection we focus on analytically determining the optimal power allocation in the high SNR regime.

### B. Optimal Power Allocation in High SNR Regime

In the high SNR regime, i.e., as $\rho_u \to \infty$ and $\rho_v \to \infty$, following (70), $R_{u1} + R_{u2}$ can be further approximated as

$$R_{u1} + R_{u2} = \log_2 \left( 1 + \phi_2 N \beta_2 \rho_u + \phi_2 (1 - \beta_2) \rho_v \right) + \log_2 \left( 1 + \frac{\phi_1 \rho_u + \phi_1 (N - 1) \beta_1 \rho_u}{\phi_2 \rho_u + 1} \right)$$

$$= \log_2 (c_2 \phi_2) + \log_2 \left( \frac{\phi_1 ((N - 1) \beta_1 + 1) \rho_u}{\phi_2 \rho_u + 1} \right)$$

$$= \log_2 (((N - 1) \beta_1 + 1) c_2 \phi_1). \quad (50)$$

Noting $\phi_1 + \phi_2 + \phi_e = 1$ and following (50), for given $\beta_1$ and $\beta_2$ the optimization problem $P2$ can be rewritten as

$\textbf{P6 :}$

$$\max_{\phi_1, \phi_2} \bar{R}_s(\beta_1, \beta_2) = \max_{\phi_1, \phi_2} \left[ \log_2 \left( \phi_1 (1 + (1 - \phi_1 - \phi_2) \rho_e)^K \right) + \log_2 \left( \frac{\phi_1 (N - 1) \beta_1 + 1) \rho_u}{\phi_2 \rho_u + 1} \right) \right] \quad (51)$$

$$\text{s.t. } \phi_1 \geq \frac{1}{c_1} (2^{Q_1} - 1)(1 + \phi_2 \rho_u), \quad (52)$$

$$\phi_2 \geq \frac{1}{c_2} (2^{Q_2} - 1). \quad (53)$$

The solution to the optimization problem $P6$ is presented in the following theorem.

**Theorem 2:** In the high SNR regime, i.e., as $\rho_u \to \infty$ and $\rho_v \to \infty$, the optimal power allocation coefficients, which are solutions to the optimization problem $P6$, are derived as

$$\phi_1^* = \begin{cases} 
1 + \rho_e - \gamma_1 \rho_e \leq 1 - \gamma_1, \\
1 - \gamma_1, \quad \text{when } \gamma_1 \leq 1 + \rho_e \gamma_1 \rho_e \leq 1 - \gamma_1, \\
\gamma_0, \quad \text{when } \gamma_0 > 1 - \gamma_1, \quad (54)
\end{cases}$$

$$\phi_2^* = \gamma_1, \quad (55)$$

$$\phi_e^* = 1 - \phi_1^* - \phi_2^*, \quad (56)$$

where $\gamma_0 = \frac{1}{c_1} (2^{Q_1} - 1)(1 + \gamma_1 \rho_u)$ and $\gamma_1 = \mu_1$.

**Proof:** As per (51), the objective function in $P6$, i.e., the SSR $\bar{R}_s(\beta_1, \beta_2)$, monotonically decreases with $\phi_2$. Noting the constraints given in (52) and (53), we conclude that the optimal value of $\phi_2$ is the one that guarantees the equality in (53), since decreasing $\phi_2$ makes the constraint given in (52) be guaranteed more easily. After obtaining the optimal value of $\phi_2$ (i.e., $\phi_2^* = \gamma_1$), the optimization problem $P6$ can be rewritten as

$$P7 : \max_{\phi_1} \bar{R}_s(\beta_1, \beta_2) = \max_{\phi_1} \left[ \log_2 \left( \phi_1 (1 + (1 - \phi_1 - \gamma_1) \rho_e)^K \right) + \log_2 \left( ((N - 1) \beta_1 + 1) c_2 \right) - K \log_2 (1 + \rho_e) \right] \quad (57)$$

$$\text{s.t. } \gamma_0 \leq \phi_1 \leq 1 - \gamma_1. \quad (58)$$
where the constraint given in (58) comes from (52), (53), and the consideration of \( \phi_1 + \phi_2 + \phi_c = 1 \) and \( 0 \leq \phi_c \).

In the following, we first maximize the objective function in \( P7 \) given in (57) without considering the constraint of (58), which is presented in the following lemma.

**Lemma 4:** The term of \( \phi_1 (1 + (1 - \phi_1 - \gamma_1)\rho_c)K \) in the objective function of \( P7 \) (i.e., the first term given in (57)) is maximized over \( \phi_1 \) when \( \phi_1 = \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \).

**Proof:** The proof is presented in Appendix C. \( \square \)

Based on (57), we find that the second and third terms in the objective function of \( P7 \) are not functions of \( \phi_1 \). Since \( \log x \) is an increasing function of \( x \), the value of \( \log x \) is maximized when \( x \) is maximized. As such, in order to maximize the objective function in \( P7 \) (i.e., (57)) without considering the constraint of (58), following Lemma 4, we have \( \phi_1 = \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \).

Now, we consider the constraint of (58) in \( P7 \). Specifically, if \( \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \) satisfies the constraints given in (58), we can directly conclude \( \phi_1^* = \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \). Otherwise, we have the following two cases, which directly follow from the proof of Lemma 4. Specifically, if \( \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} > 1 - \gamma_1 \), we have \( \phi_1^* = 1 - \gamma_1 \). Otherwise, we have \( \phi_1^* = \gamma_0 \). Finally, we can obtain the optimal power allocation coefficient for AN as \( \phi_c^* = 1 - \phi_1^* - \phi_2^* \). This completes the proof of Theorem 2.

Following Theorem 2, we note that \( \phi_2^* P \) is a fixed value regardless of the total transmit power \( P \) in the high-SNR regime, which is given by

\[
\phi_2^* P = \frac{\sigma_2^2}{(1 + (N-1)\beta_2)\sigma_2^2}(2\rho_2 - 1).
\]

This indicates that for a downlink MISO-M NOMA system, the optimal power allocation policy for maximizing the SSR is to use a fixed transmit power to User 2 in order to guarantee the equality in its QoS constraint and then allocate the remaining transmit power \( (P - P_{\text{min}}) \) to User 1 or transmitting AN signals. This is different from the conclusion drawn in [28], which is that the extra transmit power is still allocated to User 2. As per (59), we also note that the transmit power allocated to User 2 (i.e., \( \phi_2^* P \)) decreases with \( N \), which is not a function of \( K \).

Following Theorem 2, for \( \gamma_0 \leq \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \leq 1 - \gamma_1 \), \( \phi_c^* \) increases with \( K \) or \( \rho_c \), while \( \phi_1^* \) decreases with \( K \) or \( \rho_c \). This indicates that for a fixed total transmit power, we need to allocate more transmit power to AN when Eve’s channel quality becomes higher.

For \( \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} > 1 - \gamma_1 \), we have \( \phi_c^* = 0 \) as per Theorem 2, which indicates that under some specific conditions it is not necessary to transmit AN. We note that the value of \( \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \) increases when \( K \) or \( \rho_c \) decreases. This indicates that as Eve’s channel quality becomes lower, the probability of the BS having to transmit AN decreases.

Following Theorem 2, for \( \gamma_0 > \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \) in the optimal power allocation we find that the transmit power allocated to User 1 and User 2 only guarantees the equality in their QoS constraints and then all the remaining transmit power is allocated to AN. The probability of this case increases with \( K \) and \( \rho_c \), for which Eve is a very strong eavesdropper.

### C. Optimization of Beamforming Parameters \( \beta_1 \) and \( \beta_2 \)

So far, we have presented the optimization of the power allocation coefficients for given beamforming parameters \( \beta_1 \) and \( \beta_2 \). In this subsection, we discuss the optimization framework of \( \beta_1 \) and \( \beta_2 \).

We note that Lemma 1 determines a one-to-one relationship between \( \beta_1 \) and \( \beta_2 \). Specifically, following (8) and considering \( N \to \infty \), we have

\[
\gamma_1^{\text{opt}}(\beta_1, \beta_2) = \frac{\phi_1 (1 + \beta_1)\rho_c - \phi_1\beta_1\rho_c}{\phi_2\rho_2 + 1}.
\]

Following Theorem 2, we note that \( \phi_2^* P \) is a fixed value regardless of the total transmit power \( P \) in the high-SNR regime, which is given by

\[
\phi_2^* P = \frac{\sigma_2^2}{(1 + (N-1)\beta_2)\sigma_2^2}(2\rho_2 - 1).
\]

This indicates that for a downlink MISO-M NOMA system, the optimal power allocation policy for maximizing the SSR is to use a fixed transmit power to User 2 in order to guarantee the equality in its QoS constraint and then allocate the remaining transmit power \( (P - P_{\text{min}}) \) to User 1 or transmitting AN signals. This is different from the conclusion drawn in [28], which is that the extra transmit power is still allocated to User 2. As per (59), we also note that the transmit power allocated to User 2 (i.e., \( \phi_2^* P \)) decreases with \( N \), which is not a function of \( K \).

Following Theorem 2, for \( \gamma_0 \leq \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \leq 1 - \gamma_1 \), \( \phi_c^* \) increases with \( K \) or \( \rho_c \), while \( \phi_1^* \) decreases with \( K \) or \( \rho_c \). This indicates that for a fixed total transmit power, we need to allocate more transmit power to AN when Eve’s channel quality becomes higher.

For \( \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} > 1 - \gamma_1 \), we have \( \phi_c^* = 0 \) as per Theorem 2, which indicates that under some specific conditions it is not necessary to transmit AN. We note that the value of \( \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \) increases when \( K \) or \( \rho_c \) decreases. This indicates that as Eve’s channel quality becomes lower, the probability of the BS having to transmit AN decreases.

Following Theorem 2, for \( \gamma_0 > \frac{1 + \rho_c - \gamma_1}{(K+1)\rho_c} \) in the optimal power allocation we find that the transmit power allocated to User 1 and User 2 only guarantees the equality in their QoS constraints and then all the remaining transmit power is allocated to AN. The probability of this case increases with \( K \) and \( \rho_c \), for which Eve is a very strong eavesdropper.
Then, the optimization problem \( \text{P8} \) can be rewritten as

\[
\begin{align*}
\text{P9} : \quad & \max_{\beta_1, \beta_2} \hat{R}_s(\beta_1, \beta_2) \\
\text{s.t.} \quad & \beta_2 = \frac{1}{1 + N\beta_1} (1 - \beta_1), \quad (67) \\
& 0 \leq \beta_m \leq 1, \quad m \in \{1, 2\}, \quad (68)
\end{align*}
\]

which is identical to the optimization problem of \( \text{P2} \) in the high SNR regime. We note that the optimization problem \( \text{P9} \) can be efficiently solved by a one-dimensional numerical search method, since as detailed in Theorem 2 the optimal power allocation can be determined in closed form in the high SNR regime, in which the one-to-one relationship given in (67) is also independent of the power allocation coefficients. As confirmed in our following numerical results, the achieved solution to \( \text{P9} \) is very close to the solution to \( \text{P8} \) and their resultant maximum SSRs are very similar to each other. This indicates that our proposed beamforming design can be optimized efficiently by a one-dimensional numerical search.

**Remark 1:** Based on the analysis in Section IV-A, IV-B, and IV-C, we note that the proposed NOMA-HB-AN scheme requires a three-dimensional numerical search in the arbitrary SNR regime to obtain the optimal power allocation coefficients (i.e., \( \phi_1, \phi_2, \) and \( \phi_3 \)) and beamforming parameters (i.e., \( \beta_1 \) and \( \beta_2 \)), while requires a one-dimensional numerical search in high SNR regime to obtain the beamforming parameters but determines the optimal power allocation in closed form. As such, the proposed NOMA-HB-AN scheme in the high SNR regime has a much lower complexity relative to that in the arbitrary SNR regime. Furthermore, the examination results in Section V indicate that the asymptotic solution to the optimization of power allocation coefficients and governing parameters in the high SNR regime can achieve the similar secrecy performance as the exact solution in the general SNR regime. As such, the asymptotic solution in the high SNR regime can be viewed as a generic and near-optimal strategy.

## V. Numerical Results

In this section, we provide numerical results to examine the secrecy performances of the proposed NOMA-HB-AN scheme relative to two benchmark schemes. The first benchmark scheme is named as the NOMA-HB scheme, in which the beamforming design is the same as the proposed NOMA-HB-AN scheme, but no AN is transmitted by the BS. The second benchmark scheme is named as the NOMA-h2-AN scheme, which was proposed in [34]. In the NOMA-h2-AN scheme, the beamforming vectors are set as \( \mathbf{v}_1 = \mathbf{v}_2 = \mathbf{h}_2/\|\mathbf{h}_2\| \) and the AN signals are transmitted by the BS simultaneously. In this section, we set \( \rho_{su} = 1.2\rho_{su} \).

In Fig. 2, we plot the effective SSRs achieved by the NOMA-HB-AN, NOMA-HB, and NOMA-h2-AN schemes with the optimal power allocation versus the number of antennas at base station (i.e., \( N \)). We note that the effective SSRs is the sum of the effective secrecy rates at the users, where the effective secrecy rate at each user is the production of the corresponding transmission rate and the secrecy probability (which is one minus the secrecy outage probability). In this figure, we adopted reasonable transmission rates (may not be the optimal ones) at the two users and applied the developed NOMA-HB-AN scheme achieved under the assumption of a sufficiently large number of transmit antennas at the BS. In this figure, the effective SSRs are obtained by performing Monte Carlo simulations over \( 10^5 \) different channel realizations. Furthermore, from this figure we observe that the effective SSRs increase as the number of antennas at the BS increases. This is due to the fact that the extra DOF offered by additional antennas provides a more precise information beamforming that leads to a higher SINR/SNR at User 1 and User 2, while limiting the achievable data rate at the eavesdropper. In this figure, we also observe that the proposed NOMA-HB-AN scheme outperforms the NOMA-h2-AN scheme. This demonstrates the effectiveness of the proposed beamforming design, which can balance the useful signal strength and the interference strength in order to improve the secrecy performance of the considered NOMA system. We note that in this figure we set \( \beta_1 = 0.05 \) and \( \beta_2 = 0.9 \), which means that the performance gain of the proposed scheme over the NOMA-h2-AN scheme can be further improved by jointly optimizing \( \beta_1 \) and \( \beta_2 \). Finally, we observe that the proposed NOMA-HB-AN scheme outperforms the NOMA-HB scheme in terms of achieving a significantly higher effective SSR, which shows the benefits of using AN-aided transmission schemes in enhancing physical layer security of NOMA systems.
In Fig. 3, we plot the maximum SSRs of the NOMA-HB-AN, NOMA-HB, and NOMA-h_2-AN schemes versus the number of antennas at Eve (i.e., \(K\)). In this figure, the power allocation coefficients together with the values of \(\beta_1\) and \(\beta_2\) in the NOMA-HB-AN and NOMA-HB schemes have been optimized based on our conducted analysis in Section III and Section IV (e.g., Proposition 1, Theorem 1). The power allocation coefficients in the NOMA-h_2-AN scheme have been optimized based on the analysis presented in [34]. We first observe that the analytical curve of the NOMA-HB-AN scheme accurately match the simulated one for different values of \(N\), which confirms the high accuracy of the approximation adopted in our Proposition 1. Again, in this figure, we first observe that the proposed NOMA-HB-AN scheme significantly outperforms the two benchmark schemes, which demonstrates the superiority of jointly using the proposed beamforming design and AN-aided transmission strategy. As expected, we observe that the achieved maximum SSRs decrease with the number of antennas at the eavesdropper (i.e., \(K\)). This is due to the fact that the downlink NOMA transmission becomes more vulnerable to eavesdropping attack when the eavesdropper has more receiving antennas in the considered system. Therefore, more DOF (e.g., more transmit antennas at the BS) should be utilized to achieve the same level of secrecy as \(K\) increases. We further observe that the performance gain of the proposed scheme over the NOMA-h_2-AN scheme increases with \(K\). This indicates that the advantage of the proposed scheme relative to the NOMA-h_2-AN scheme becomes more dominant as \(K\) increases, which shows the effectiveness of the proposed beamforming design increases with \(K\), while the NOMA-h_2-AN scheme employ the fixed beamforming (i.e., \(v_1 = v_2 = \hat{h}_2/\|\hat{h}_2\|\)) causes severe information leakage to Eve and severe interference to User 1 with \(K\).

We further examine the impact of the proposed NOMA-HB-AN scheme on the secrecy outage probability when the number of antennas at the BS is finite. In Fig. 4, we plot the secrecy outage probabilities achieved by the NOMA-HB-AN, NOMA-HB, and NOMA-h_2-AN schemes with the optimal power allocation versus the number of antennas at BS (i.e., \(N\)). In this figure, we set the secrecy transmission rates to \(2648\) BPCU and \(5.5\) BPCU. In this figure, the secrecy outage probabilities are obtained by performing Monte Carlo simulations over \(10^6\) different channel realizations. From this figure we observe that the secrecy outage probability decreases as when \(N\) increases. This is due to the fact that the extra DOF offered by additional antennas provides a more precise information beamforming that leads to a higher SNR or SNR at User 1 and User 2, while limiting the achievable data rate at the eavesdropper. In this figure, we also observe that the secrecy outage probability advantage of the NOMA-HB-AN scheme over the NOMA-h_2-AN scheme becomes more prominent when \(N\) increases. This again demonstrates the effectiveness of the proposed beamforming design and these observations are in accordance with the observations made from the analytical results. Furthermore, we observe that the proposed NOMA-HB-AN scheme outperforms the NOMA-HB scheme in terms of achieving a significantly lower secrecy outage probability, illustrating the security benefits.
For the term \( R_{e1} + R_{e2} \) in (13), which can be further expressed as (71), shown at the top of this page, where \( \pi_e = \frac{P}{\sigma_n^2} \). In order to simplify (69), we present the following lemma (i.e., Lemma 5) to facilitate our proof.

**Lemma 5:** As \( N \to \infty \) with \( N \gg K \), we have

\[
\det \left( I_K + \phi_1 \pi_c H_e v_1 (H_e v_1)^H + \phi_2 \pi_e H_e v_2 (H_e v_2)^H + \phi_e \pi_e H_e V_N (H_e V_N)^H \right)
\]
\[
\leq \prod_{k=1}^{K} \left( 1 + \phi_1 \pi_c [H_e v_1 (H_e v_1)^H]_{kk} + \phi_2 \pi_e [H_e v_2 (H_e v_2)^H]_{kk} + \frac{(1 - \phi_1 - \phi_2 \pi_e)}{N-2} [H_e V_N (H_e V_N)^H]_{kk} \right)
\]
\[
\approx (1 + \phi_1 \rho_e + \phi_2 \rho_e + (1 - \phi_1 - \phi_2 \rho_e)^K
\]
\[
= (1 + \rho_e)^K,
\]
and

\[
\det \left( I_K + \phi_1 \pi_c H_e V_N (H_e V_N)^H \right)
\]
\[
\approx \prod_{k=1}^{K} \left( 1 + \frac{\phi_1 \pi_c}{N-2} [H_e V_N (H_e V_N)^H]_{kk} \right)
\]
\[
= (1 + \rho_e)^K.
\] (73)

The proof of [44, Lemma 5] and thus it is omitted here. Based on (70), (71), and Lemma 5, we can obtain (25), which completes the proof of Proposition 1.

**APPENDIX B**

**PROOF OF LEMMA 3**

Substituting (39) into (37), we have

\[
F(\phi_1, \phi_2) = \log_2 \left( 1 + \frac{c_1 (1 - \phi_2 - \phi_e)}{1 + \phi_2 \rho_{u1}} \right) (1 + c_2 \phi_2)
\]
\[
= G(\phi_2).
\] (74)

From (74), we can easily obtain the first derivative of \( G(\phi_2) \) with respect to \( \phi_2 \) as \( G(\phi_2)' \), which is given by

\[
G(\phi_2)' = \frac{\partial G(\phi_2)}{\partial \phi_2} = \frac{1}{\ln(2)} \left( c_2 \rho_{u1} \phi_2^2 + 2 c_2 \phi_2 + c_4 \right),
\] (75)

where

\[
c_3 = \rho_{u1} - 1,
\]
\[
c_4 = (1 + (1 - \phi_2)) (c_2 - \rho_{u1}) + \rho_{u1} - 1.
\] (76)

Furthermore, we can obtain the second derivative of \( G(\phi_2) \) with respect to \( \phi_2 \) as \( G(\phi_2)'' \), which is given by

\[
G(\phi_2)'' = \frac{\partial G(\phi_2)'}{\partial \phi_2} = \frac{1}{\ln(2)} \left( c_1 (2 - \phi_2) ((N-1) \beta_2 \rho_{u2} + \rho_{u2} - \rho_{u1}) \right) < 0,
\] (78)

due to the facts \( \phi_e < 1 \) and \( \rho_{u1} \leq \rho_{u2} \). Following (78), we find that \( G(\phi_2) \) is a concave function of \( \phi_2 \), which completes the proof of Lemma 3.

**APPENDIX C**

**PROOF OF LEMMA 4**

Following (57), we derive the first derivative of \( \phi_1 (1 + (1 - \phi_1 - \gamma_1) \rho_e)^K \) with respect to \( \phi_1 \) as

\[
Q(\phi_1)' \triangleq \frac{\partial Q(\phi_1)}{\partial \phi_1} = ((1 + (1 - \phi_1 - \gamma_1) \rho_e)^{-K-1} \times ((1 + (1 - \phi_1 - \gamma_1) \rho_e) - K \phi_1 \rho_e). \] (79)

Following (79) and noting \( 1 - \phi_1 - \gamma_1 \geq 0 \), we have

\[
\phi_1 = \phi_1^+ \triangleq 1 + \rho_e - \gamma_1 \rho_e
\]
\[
(K + 1) \rho_e,
\] (80)

in order to guarantee \( Q(\phi_1)' = 0 \). As per (79), we can find that \( Q(\phi_1)' > 0 \) for \( \phi_1 < \phi_1^+ \), which indicates that...
the function of $Q(\phi_1)$ is a monotonically increasing function of $\phi_1$ when $\phi_1 > \phi_1'$. We also find that $Q(\phi_1') < 0$ for $\phi_1 > \phi_1'$, which shows that the function of $Q(\phi_1)$ is a monotonically decreasing function of $\phi_1$ when $\phi_1 > \phi_1'$. As such, we can conclude that $Q(\phi_1')$ is maximized when $Q(\phi_1') = 0$, i.e., when (80) is guaranteed. This completes the proof of Lemma 4.

References

[1] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, “System-level performance evaluation of downlink non-orthogonal multiple access (NOMA),” in Proc. IEEE PIMRC, London, U.K., Sep. 2013, pp. 611–615.

[2] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, “Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends,” IEEE Commun. Mag., vol. 53, no. 9, pp. 74–81, Sep. 2015.

[3] 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.

[4] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, “Nonorthogonal multiple access for 5G and beyond,” Proc. IEEE, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[5] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, “Optimal joint power and subcarrier allocation for full-duplex multi-carrier non-orthogonal multiple access systems,” IEEE Trans. Wireless Commun., vol. 65, no. 3, pp. 1077–1091, Mar. 2017.

[6] Y. Feng, S. Yan, C. Liu, Z. Yang, and N. Yang, “Two-stage relay selection for enhancing physical layer security in non-orthogonal multiple access,” IEEE Trans. Inf. Forensics Security, vol. 14, no. 6, pp. 1670–1683, Jun. 2019.

[7] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, “Optimal joint power and subcarrier allocation for MC-NOMA systems,” in Proc. IEEE Global Commun. Conf., Washington, DC, USA, Dec. 2016, pp. 1–6.

[8] A. Benjebboura, A. Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, “System-level performance of downlink NOMA for future LTE enhancements,” in Proc. IEEE GLOBECOM, Dec. 2013, pp. 66–70.

[9] J. Choi, “On the power allocation for MIMO-NOMA systems with layered transmissions,” IEEE Trans. Wireless Commun., vol. 15, no. 5, pp. 3226–3237, May 2016.

[10] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, “Short-packet downlink transmission with non-orthogonal multiple access,” IEEE Trans. Wireless Commun., vol. 17, no. 7, pp. 4550–4564, Jul. 2018.

[11] X. Sun et al., “Joint beamforming and power allocation in downlink NOMA multiuser MIMO networks,” IEEE Trans. Wireless Commun., vol. 17, no. 8, pp. 5367–5381, Aug. 2018.

[12] H. Wang, R. Zhang, R. Song, and S.-H. Leung, “A novel power minimization precoding scheme for MIMO-NOMA uplink systems,” IEEE Commun. Lett., vol. 22, no. 5, pp. 1106–1109, May 2018.

[13] H. Wang, S.-H. Leung, and R. Song, “Precoding design for two-cell MIMO-NOMA uplink with CoMP reception,” IEEE Commun. Lett., vol. 22, no. 12, pp. 2607–2610, Dec. 2018.

[14] T. Cover, “Broadcast channels,” IEEE Trans. Inf. Theory, vol. 18, no. 1, pp. 2–14, Jan. 1972.

[15] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, “V-BLAST: An architecture for realizing very high data rates over the air,” IEEE Electron. Lett., vol. 33, no. 12, pp. 1664–1665, Jun. 1997.

[16] Z. Ding, F. Adachi, and H. V. Poor, “The application of MIMO to non-orthogonal multiple access,” IEEE Trans. Wireless Commun., vol. 15, no. 1, pp. 537–552, Jan. 2016.

[17] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, “A survey on wireless security: Technical challenges, recent advances, and future trends,” Proc. IEEE, vol. 104, no. 9, pp. 1727–1765, Sep. 2016.

[18] T.-X. Zheng, H. Wang, J. Yuan, D. Towsley, and M. H. Lee, “Multi-antenna transmission with artificial noise against randomly distributed eavesdroppers,” IEEE Trans. Commun., vol. 63, no. 11, pp. 4347–4362, Nov. 2015.
Youhong Feng (S’16–M’18) received the B.S. and M.S. degrees in information engineering from Chang’an University, Xi’an, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Nanjing University of Posts and Telecommunications, China, in 2010. From 2010 to 2011, he was a Visiting Student with the National University of Singapore. From 2011 to 2014, he was a Senior Lecturer at South China University of Technology. His research interests include wireless communications, ultra-reliable low latency communications, and physical layer security.

Shihao Yan (M’15) received the B.S. degree in communication engineering and the M.S. degree in communication and information systems from Shandong University, Jinan, China, in 2009 and 2012, respectively, and the Ph.D. degree in electrical engineering from the University of New South Wales, Sydney, Australia, in 2015. From 2015 to 2017, he was a Post-Doctoral Research Fellow with the Research School of Engineering, The Australian National University, Canberra, Australia. He is currently an Associate Professor with the College of Physics and Electronic Information Engineering, Anhui Normal University. His research interests include cooperative communications, energy-efficient communications, ultra-reliable low latency communications, and physical layer security.

Zhen Yang received the B.Eng. degrees in electrical engineering from the Nanjing University of Posts and Telecommunications, China, in 1983 and 1988, respectively, and the Ph.D. degree in electrical engineering from Shanghai Jiao Tong University, China, in 1999. He was a Lecturer with the Nanjing University of Posts and Telecommunications in 1983, where he was promoted as an Associate Professor in 1995 and has been a Full Professor since 2000. He was a Visiting Scholar with Bremen University, Germany, from 1992 to 1993, and an Exchange Scholar with Maryland University, USA, in 2003. He has published more than 200 papers in academic journals and conferences. His research interests include various aspects of signal processing and communication, such as communication systems and networks, cognitive radio, spectrum sensing, speech and audio processing, compressive sensing, and wireless communication.

He is a member of the Editorial Boards of several other journals, such as the Chinese Journal of Electronics, China Communications, and Data Collection and Processing, the Chair of Asian Pacific Communication Conference (APCC) Steering Committee from 2013 to 2014. He serves as a Vice Chairman and a Fellow of the Chinese Institute of Communications, a Chairman of the Jiangsu Institute of Internets, and the Vice Director of the Editorial Board of the Journal of Communications.

FENG et al.: BEAMFORMING DESIGN AND POWER ALLOCATION FOR SECURE TRANSMISSION WITH NOMA

[43] S.-H. Tsai and H. Poor, “Power allocation for artificial-noise secure MIMO precoding systems,” IEEE Trans. Signal Process., vol. 62, no. 13, pp. 3479–3493, Jul. 2014.

[44] N. Li, X. Tao, H. Wu, J. Xu, and Q. Cui, “Large-system analysis of artificial-noise-assisted communication in the multiuser downlink: Ergodic secrecy sum rate and optimal power allocation,” IEEE Trans. Veh. Technol., vol. 65, no. 9, pp. 7026–7030, Sep. 2016.

[45] N. Li, X. Tao, and J. Xu, “Artificial noise assisted communication in the multiuser downlink: Optimal power allocation,” IEEE Commun. Lett., vol. 19, no. 2, pp. 295–298, Feb. 2015.

[46] Z. Ding et al., “Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions,” IEEE Trans. Veh. Technol., vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[47] H. Lei et al., “On secure NOMA systems with transmit antenna selection schemes,” IEEE Access, vol. 5, pp. 17450–17464, 2017.

[48] Y. Li, M. Jiang, Q. Zhang, Q. Li, and J. Qin, “Secure beamforming in downlink MISO nonorthogonal multiple access systems,” IEEE Trans. Veh. Technol., vol. 66, no. 8, pp. 7563–7567, Aug. 2017.

[49] L. Lv, J. Chen, Q. Ni, and Z. Ding, “Design of cooperative orthogonal multicast cognitive multiple access for 5G systems: User scheduling and performance analysis,” IEEE Trans. Commun., vol. 65, no. 6, pp. 2641–2656, Jun. 2017.

[50] M. F. Hanif, Z. Ding, T. Rattaarajah, and G. K. Karagiannidis, “A minimization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems,” IEEE Trans. Signal Process., vol. 64, no. 1, pp. 76–88, Jan. 2015.

[51] Y. Sun, D. W. K. Ng, J. Zhu, and R. Schober, “Robust and secure resource allocation for full-duplex MISO multi-carrier NOMA systems,” IEEE Trans. Veh. Technol., vol. 66, no. 9, pp. 4119–4137, Sep. 2018.

[52] D. Zhang, Y. Liu, Z. Ding, Z. Zhou, Nallanathan, and T. Sato, “Performance analysis of non-regenerative massive-MIMO-NOMA relay systems for 5G,” IEEE Trans. Commun., vol. 65, no. 11, pp. 4777–4790, Nov. 2017.

Nan Yang (S’09–M’11–SM’18) received the B.S. degree in electronics from China Agricultural University in 2005, and the M.S. and Ph.D. degrees in electronic engineering from the Beijing Institute of Technology, in 2007 and 2011, respectively. He was a Post-Doctoral Research Fellow with the University of New South Wales Sydney, from 2012 to 2014 and with the Commonwealth Scientific and Industrial Research Organization from 2010 to 2012. He has been with the Research School of Engineering, Australian National University, since 2014, where he is currently a Future Engineering Research Leadership Fellow and a Senior Lecturer. His general research interests include massive multi-antenna systems, millimeter wave and terahertz communications, ultra-reliable low latency communications, cyber-physical security, and molecular communications. He was a recipient of the Top Editor Award from the Transactions on Emerging Telecommunications Technologies in 2017, the Exemplary Reviewer Award of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2015 and 2016, the Top Reviewer Award from the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in 2015, the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award and the Exemplary Reviewer Award of the IEEE WIRELESS COMMUNICATIONS LETTERS in 2014, and the Exemplary Reviewer Award of the IEEE COMMUNICATIONS LETTERS in 2012 and 2013. He was also a Co-Recipient of the Best Paper Awards from the IEEE GLOBECOM 2016 and the IEEE VTC 2013-Spring. He is currently serving on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and Transactions on Emerging Telecommunications Technologies.

Jinhong Yuan (M’02–SM’11–F’16) received the B.E. and Ph.D. degrees in electronics engineering from the Beijing Institute of Technology, Beijing, China, in 1991 and 1997, respectively. From 1997 to 1999, he was a Research Fellow with the School of Electrical Engineering, University of Sydney, Sydney, Australia. In 2000, he joined the School of Electrical Engineering and Telecommunications, University of New South Wales Sydney, Sydney, Australia, where he is currently a Professor and the Head of the Telecommunication Group. He has published two books, five book chapters, over 500 papers in telecommunications journals and conference proceedings, and 50 industrial reports. He is a co-inventor of one patent on MIMO systems and two patents on low-density-parity-check codes. His current research interests include error control coding and information theory, communication theory, and wireless communications. He has co-authored four Best Paper Awards and one Best Poster Award, including the Best Paper Award from the IEEE International Conference on Communications, Kansas City, USA, in 2018, the Best Paper Award from IEEE WIRELESS COMMUNICATIONS AND NETWORKING CONFERENCE, Cancun, Mexico, in 2011, and the Best Paper Award from the IEEE International Symposium on Wireless Communications Systems, Trondheim, Norway, in 2007. He served as the IEEE NSW Chapter Chair for Joint Communications/Signal Processing/Ocean Engineering Chapter from 2011 to 2014. He served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 2012 to 2017. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.