This article has been accepted for publication in IEEE Wireless Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LWC.2022.3197633

Beyond Non-Orthogonal Multiple Access: New Role of Constructive Interference

Zhongxiang Wei, *Member, IEEE,* Christos Masouros, *Senior Member, IEEE,* Ping Wang, *Member, IEEE,* Xu Zhu, *Senior Member, IEEE,* Haiyong Zeng, *Member, IEEE,* Hong Tang, and Yan Zheng

***Abstract*—In this paper, we introduce a novel framework of constructive non-orthogonal multiple access (NOMA) transmission, which provides the merit of interference utilization and breaks through the constructive interference (CI)'s limitation on multiuser (MU) access capability. With dedicated synthetic successive coding and hybrid MU access designs, a novel constructive NOMA (CNOMA) precoder is proposed, which is particularly suitable for the scenario where users have heterogeneous throughput requirements. Explicitly, it makes the composite interference always beneficial to the users having high throughput requirement, while accommodating another sets of users under their subscribed reception-quality requirement. Finally, a number of fundamental properties of the CNOMA design is revealed, such as the tradeoff between utilization of MU interference and improvement of MU access capability. Simulation demonstrates that the proposed CNOMA precoder significantly outperforms the classic CI and minimum-mean-square-error precoders in throughput performance, and meanwhile obtains high access capability close to classic NOMA designs.**

***Index Terms*—constructive NOMA design, interference exploitation, multiuser access capability, precoding.**

I. Introduction

In last decades, a series of precoders has been proposed for multiuser (MU) multiple-input and multiple-output (MIMO) systems. The capacity-approaching non-linear dirty-paper coding and vector perturbation precoders generally incur high complexity, due to the inclusion of the sophisticated sphere-search algorithms. Linear zero-forcing (ZF) and minimum mean squared error (MMSE) precoders were introduced to strike the balance between performance and complexity [1]. Also, optimization-based precoding, such as signal-to-interference-plus-noise ratio (SINR) balancing [2] and power minimization precoders [3], were extensively researched. The principle of the precoders above is to cancel MU interference (MUI) by exploiting the spatial multiplexing of MIMO channels [4]. They generally require that the number of transmit-antennas is no less than that of the users. Otherwise, these precoders either are infeasible due to the rank-deficiency property of the precoder design (e.g., ZF and optimization-based precoders), or lead to significantly degraded performance due to the low degrees-of-freedom (DoF) (e.g., MMSE precoder).

Non-orthogonal-multiple-access (NOMA) has received considerable attention due to its high access capability. The key of NOMA is that of allowing multiple users to use the same frequency-, time-, space- or code-domain resources [5] [6] [7]. Recently, the concept of NOMA was extended to rate-splitting [8], phase rotation-based NOMA [9] and semi-orthogonal multiple access [10]. The integration of NOMA and MIMO was investigated in [11] [12], where the ergodic capacity and the impact of user pairing were analyzed. Since precoding was not involved, the number of receive-antenna of each user should be no less than that of transmit-antennas of the base station (BS) [13]. A ZF NOMA precoder was proposed to exploit the channel spatiality [13], where the DoF of the precoder is strictly constrained for suppressing the inter-group interference. In [8], the authors studied the DoF of NOMA against MU linear precoding as well as rate-splitting multiple access. Hence, the existing NOMA MIMO designs still try to provide orthogonal spatial transmission among NOMA groups to avoid MUI. Based on the concept of constructive interference (CI) [14], it is possible to exploit the correlation between the symbols, so that MUI being aligned to the signal of interest at each receiver acts as a constructive element [15]. If the MUI can be utilized, it will potentially contribute to system performance, and also the DoF in precoding design can be relaxed over the existing NOMA MIMO designs.

The orchestration of CI and NOMA, however, is not straightforward. It involves hybrid MU access framework, and dedicated precoding/combiner designs, which are particularly challenging in a dense communication scenario. It is because CI requires that the number of users is no larger than that of the transmit-antenna, and also the signal superposition mechanism of NOMA further complicates the MUI exploitation. In this work, we present a first attempt to design constructive NOMA transmission. Our contributions are summarized as follows.

1) It is the first contribution demonstrating constructive NOMA framework for MU MIMO communications. This framework utilizes MUI as a beneficial element for improving receive-performance, and meanwhile breaks through the CI's limitation on MU access capability. With dedicated synthetic successive coding and hybrid

Zhongxiang Wei and Ping Wang are with the College of Electronic and Information Engineering, at Tongji University, Shanghai, China. Email: {z_wei, pwang}@tongji.edu.cn

Christos Masouros is with the Department of Electronic and Electrical Engineering, at University College London, London, UK. Email: c.masouros@ucl.ac.uk

Xu Zhu and Haiyong Zeng are with the School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen, China. Email: {zenghaiyong, xuzhu}@stu.hit.edu.cn

Hong Tang is with Chongqing University of Posts and Telecommunication, Chongqing, China. Email: tangh@cqupt.edu.cn

Yan Zheng is with Huayu Automative System Company Limited., Email: yzheng_jd@163.com

Corresponding authors: Yan Zheng and Xu Zhu. This work is funded by the NSFC 62101384, the Chongqing Key Laboratory of Mobile Communication Technology cqupt-mct-202101, the Shenzhen S&T Program KQTD20190929172545139, and the Shenzhen Key Laboratory ZDSYS20210623091808025.

This article has been accepted for publication in IEEE Wireless Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LWC.2022.3197633

MU access designs, it endorses the ability of downlink MUI utilization, while achieving an enhanced MU access capability over the classic CI precoders.
2) Then, a novel constructive NOMA (CNOMA) precoder is proposed, which is particularly suitable for the scenario where users have heterogeneous throughput demands. In details, it provides constructive MUI for the users having high throughput requirement, and meanwhile exploits the merits of NOMA to accommodate another sets of users, under their subscribed SINR requirements. In addition, the CNOMA precoder locates the received symbol into desired regions of signal demodulation without involving semi-definite programming (SDP). Hence, the subsequent semi-definite matrix decomposition and phase equalization are not required.
3) Our work studies the fundamental tradeoff between utilizing constructive MUI and improving MU access capability. Explicitly, when accommodating more NOMA-based users, the ability of utilizing MUI as a beneficial element is inevitably reduced. Hence, the framework presents a tunable performance between the exploitation of MUI and the enhancement of MU access capability.

Notation: $|\cdot|$ denotes absolute value of a complex number or cardinality of a set. $||\cdot||$ denotes 2-norm operation. $(\cdot)^T$ and $(\cdot)^H$ denote transpose or Hermitian transpose operation.

II. System Model and Problem Formulation

In this section, system model and problem formulation are introduced, respectively.

A. System Model

Consider a dense scenario, where the number of users is larger than that of the BS's transmit-antennas. Assume the BS is equipped with N_t antennas, while each user has N_r receive-antennas. Without loss of generality, there are K_1 ($K_1 = |\mathbb{K}_1|$) users having high throughput requirement, while K_2 ($K_2 = |\mathbb{K}_2|$) users only need to access networks with relatively low throughput demand, such as requiring signalling bits.

1) Synthetic Successive Coding: Denote $\boldsymbol{a} \in \mathbb{C}^{K_1\times 1}$ and $\boldsymbol{b} \in \mathbb{C}^{K_2\times 1}$ as the desired symbol vectors for the K_1 and K_2 users by phase-shift-keying (PSK) modulation. The BS first applies successive coding and pairs the K_2 users into $\frac{K_2}{2}$ NOMA groups[1], yielding $\hat{\boldsymbol{b}} = [\rho_1 b_1 + \rho_{1'} b_{1'}, ..., \rho_{\frac{K_2}{2}} b_{\frac{K_2}{2}} + \rho_{\frac{K_2'}{2}} b_{\frac{K_2'}{2}}]^T \in \mathbb{C}^{\frac{K_2}{2}\times 1}$. ρ_m and $\rho_{m'}$ denote the power splitting factors in the m-th group, and we have $\rho_m^2 + \rho_{m'}^2 = 1$. As extensively assumed in the papers [12] [13], we consider fixed values of power splitting factors in each NOMA group. Then, synthesizing $\boldsymbol{a}$ and $\hat{\boldsymbol{b}}$ yields the transmitted symbol $\boldsymbol{s} = [\boldsymbol{a}^H, \hat{\boldsymbol{b}}^H]^H \in \mathbb{C}^{(K_1+\frac{K_2}{2})\times 1}$, where we have $K_1 + \frac{K_2}{2} = N_t$.

[1]For low-complexity successive interference cancellation (SIC), there are two users in each NOMA group, i.e, a strong user m and a weak user m', while the following can be straightforwardly extended to multiple users groups. There have been extensive works on grouping or fairness designs. Since it is not focus of our work, we refer readers to [16] [17] for details.

2) Hybrid Multiuser Access Model: Denote $\boldsymbol{W} = [\boldsymbol{w}_1, ..., \boldsymbol{w}_{N_t}] \in \mathbb{C}^{N_t\times N_t}$ as the precoder employed at the BS, with each vector representing the weight for the associated symbol in $\boldsymbol{s}$. Denote $\boldsymbol{H}_i \in \mathbb{C}^{N_r\times N_t}$ as the channel from the BS to the i-th user in $\mathbb{K}_1$. Then the received signal is calculated as $\boldsymbol{y}_i = \boldsymbol{H}_i / d_i^{\lambda/2} \boldsymbol{W}\boldsymbol{s} + \boldsymbol{n}_i$, where d_i denotes the distance between the BS and the i-th user. λ represents the path loss exponent, and $\boldsymbol{n}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_r}\sigma^2)$ denotes the complex circular Gaussian noise. With a combiner $\boldsymbol{c}_i \in \mathbb{C}^{N_r\times 1}$, the post-combined signal is given as $r_i = \boldsymbol{c}_i^H \boldsymbol{H}_i \boldsymbol{W}\boldsymbol{s}/d_i^{\lambda/2} + \boldsymbol{c}_i^H \boldsymbol{n}_i$, where the signal-to-interference-plus-noise ratio (SINR) is calculated as (1). In a similar fashion, define $\boldsymbol{G}_m \in \mathbb{C}^{N_r\times N_t}$ and d_m as the channel and communication distance from the BS to the strong user in the m-th NOMA group. Then its signal is given as $r_m = \boldsymbol{v}_m^H \boldsymbol{G}_m \boldsymbol{W}\boldsymbol{s}/d_m^{\lambda/2} + \boldsymbol{v}_m^H \boldsymbol{n}_m$, where $\boldsymbol{v}_m \in \mathbb{C}^{N_r\times 1}$ denotes the combiner applied at the strong user. In each NOMA group, the strong user first applies SIC to subtract the weak user's interference, and then decodes its own signal. For the purpose of success of SIC, it is important to guarantee that the SINR of the strong user for receiving the signal of the weak user is greater than or equal to that of the weak user for decoding its own signal. The post-SIC SINR is given as (2). On the other hand, define $\boldsymbol{G}_{m'} \in \mathbb{C}^{N_r\times N_t}$ and $d_{m'}$ as the channel and distance from the BS to the weak user in the m-th NOMA group. Define $\boldsymbol{v}_{m'} \in \mathbb{C}^{N_r\times 1}$ as the combiner applied at the weak user. Since the weak user decodes its own signal directly, its SINR is calculated by (3).

B. Problem Formulation

Under heterogeneous throughput requirements, we aim at maximizing the minimum SINR of the users having high throughput requirement, and in the meantime accommodate another set of NOMA users subject to their subscribed SINR thresholds. The optimization problem is formulated as

$$\begin{aligned}
&P1 \,:\, \underset{\boldsymbol{W},\boldsymbol{c}_i,\boldsymbol{v}_m,\boldsymbol{v}_{m'}}{\operatorname{argmax}} \quad \min \Gamma_i, \\
&(\text{C1}): K_1 + K_2 > N_t,\ (\text{C2}): \Gamma_m \geq \bar{\Gamma}_m, \text{and } \Gamma_{m'} \geq \bar{\Gamma}_m, \forall m, \\
&(\text{C3}): \Gamma_{m'm} \geq \bar{\Gamma}_{m'}, \forall m,\ (\text{C4}): ||\boldsymbol{W}\boldsymbol{s}||^2 \leq N_t p_{\max},
\end{aligned} \tag{4}$$

where (C1) denotes that we need to accommodate more users than the number of the transmit-antennas. (C2) guarantees the SINR threshold $\bar{\Gamma}_m$ of the m-th NOMA group. Constraint (C3) guarantees that in each NOMA group, the SINR of user m for receiving the signal of user m' is greater than or equal to the SINR of user m' for decoding its own signal. (C4) confines a power budget constraint $p_{\max}$. Evidently, P1 is a non-convex quadratically constrained programming. The difficulty of solving P1 lies in properly handling MUI when $K_1 + K_1 > N_t$. A possible solution would be applying the classic SDP, which however requires the number of users is no larger than that of the transmit-antenna. In the followings, we transform P1 into a tractable optimization P2, and make a good tradeoff between system performance and complexity.

III. Constructive NOMA Precoder Design

Conventional interference mitigation based precoders locate the desired signal in the proximity around the constellation

$$\Gamma_i = \frac{|\boldsymbol{c}_i^H \boldsymbol{H}_i \boldsymbol{w}_i a_i|^2/d_i^{\lambda}}{\sum_{j=1,j\neq i}^{K_1} |\boldsymbol{c}_i^H \boldsymbol{H}_i \boldsymbol{w}_j a_j|^2/d_i^{\lambda} + \sum_{m=1}^{K_2/2} |\boldsymbol{c}_i^H \boldsymbol{H}_i \boldsymbol{w}_m(\rho_m b_m + \rho_{m'} b_{m'})|^2/d_i^{\lambda} + |\boldsymbol{c}_i^H \boldsymbol{n}_i|^2}, \tag{1}$$

$$\Gamma_m = \frac{|\boldsymbol{v}_m^H \boldsymbol{G}_m \boldsymbol{w}_m \rho_m b_m|^2/d_m^{\lambda}}{\sum_{i=1}^{K_1} |\boldsymbol{v}_m^H \boldsymbol{G}_m \boldsymbol{w}_i a_i|^2/d_m^{\lambda} + \sum_{n\neq m,n=1}^{K_2/2} |\boldsymbol{v}_m^H \boldsymbol{G}_m \boldsymbol{w}_n(\rho_n b_n + \rho_{n'} b_{n'})|^2/d_m^{\lambda} + |\boldsymbol{v}_m^H \boldsymbol{n}_m|^2}, \tag{2}$$

$$\Gamma_{m'} = \frac{|\boldsymbol{v}_{m'}^H \boldsymbol{G}_{m'} \boldsymbol{w}_m \rho_{m'} b_{m'}|^2/d_{m'}^{\lambda}}{\sum_{i=1}^{K_1} |\boldsymbol{v}_{m'}^H \boldsymbol{G}_{m'} \boldsymbol{w}_i a_i|^2/d_{m'}^{\lambda} + \sum_{n\neq m,n=1}^{K_2/2} |\boldsymbol{v}_{m'}^H \boldsymbol{G}_{m'} \boldsymbol{w}_n(\rho_n b_n + \rho_{n'} b_{n'})|^2/d_{m'}^{\lambda} + |\boldsymbol{v}_{m'}^H \boldsymbol{G}_{m'} \boldsymbol{w}_m \rho_m b_m|^2/d_{m'}^{\lambda} + |\boldsymbol{v}_{m'}^H \boldsymbol{n}_{m'}|^2}, \tag{3}$$

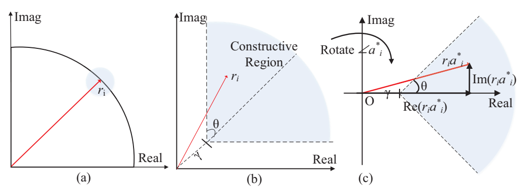

Fig. 1. The illustration of CI, where the intended symbol is $(1+i)/\sqrt{2}$.

point, as shown in Fig. 1(a). Nevertheless, since the transmitted symbols are known at the BS, we can exploit the correlation among the channels and symbols, and leverage MUI as a constructive component to improve system performance [14].

Lemma 1: Let us consider PSK modulation for illustration, nevertheless the following is also applicable to quadrature amplitude modulation [14]. For the i-th user in $\mathbb{K}_1$, write the noise-excluding received signal as $\hat{r}_i$, which can be rotated by an angle $\angle a_i^*$ (The angle depends on the conjugate of its desired symbol a_i, which is naturally known at the BS.) Then, it is projected onto real axis $\Re\{\hat{r}_i a_i^*\}$ and imaginary axis $\Im\{\hat{r}_i a_i^*\}$, respectively. Then $\hat{r}_i$ locates in a constructive region (in Fig. 1 (b)) when the inequality $|\Im\{\hat{r}_i a_i^*\}| \leq (\Re\{\hat{r}_i a_i^*\} - \sigma\sqrt{\Gamma_i})\tan\theta$ holds (in Fig. 1 (c)), where $\theta = \frac{\pi}{L}$ and L represents modulation size. □

Indicated by Lemma 1, CI exploits the knowledge of both channel and symbols, while each user only has knowledge of their own downlink channel. Due to the asymmetric information at the BS and users, the CI-based users' combiner should be channel-only dependent, such as equal gain or singular value decomposition based combiner. Then, with the combiner $\boldsymbol{c}_i$, the composite interference of the i-th user can be made beneficial as long as the following inequality is guaranteed

$$|\Im\{\boldsymbol{h}_i \boldsymbol{W} \boldsymbol{s} a_i^*\}| \leq (\Re\{\boldsymbol{h}_i \boldsymbol{W} \boldsymbol{s} a_i^*\} - \sigma\sqrt{\Gamma_i})\tan\theta, \forall i \in \mathbb{K}_1, \tag{5}$$

where $\boldsymbol{h}_i = \boldsymbol{c}_i^H \boldsymbol{H}_i d_i^{\lambda/2}$ denotes the equivalent multi-input and single-output channel of the i-th CI user.

Lemma 2: Define $\gamma_i = \sigma\sqrt{\Gamma}_i$, $\forall i \in \mathbb{K}_1$, which physically denotes the Euclidean distance in the signal constellation between the constructive region and the decision thresholds. Hence, maximizing the minimum SINR is equivalent to maximizing the minimum γ_i among the K_1 users. □

Nevertheless, if one attempts to generate constructive MUI for the NOMA users, the constraints in (5) should be imposed for all the K_2 users. As a result, it reduces to a conventional CI precoder, where its performance is significantly reduced when $K_1+K_2 > N_t$. One straightforward approach is to serve fewer users, which however reduces the MU access capability. Under the SINR requirement of (C2), we instead suppress the MUI for the NOMA-based users in $\mathbb{K}_2$. Let us first focus on $\boldsymbol{w}_i$ that is the precoder vector for the i-th CI user. Then the condition of suppressing the MUI from the K_1 users is given as

$$[\boldsymbol{G}_1^H \boldsymbol{v}_1, \boldsymbol{G}_{1'}^H \boldsymbol{v}_{1'}, ..., \boldsymbol{G}_{\frac{K_2}{2}}^H \boldsymbol{v}_{\frac{K_2}{2}}, \boldsymbol{G}_{\frac{K_2'}{2}} \boldsymbol{v}_{\frac{K_2'}{2}}]^H \boldsymbol{w}_i = \boldsymbol{0}_{K_2\times 1}, \tag{6}$$

where a non-zero $\boldsymbol{w}_i$ satisfying (6) generally exists if $K_2 \leq N_t$. However, if the dimension of the matrix can be reduced, the DoF of the precoder can be accordingly improved. In this context, we utilize the concept of signal alignment [13]. Explicitly, we manipulate the combiner of the NOMA users so that the channels of the users in the same group can be aligned into the same direction, given as $[\boldsymbol{G}_m^H, -\boldsymbol{G}_{m'}^H][\boldsymbol{v}_m^H, \boldsymbol{v}_{m'}^H]^H = \boldsymbol{0}_{N_t\times 1}$, $\forall m$. Denote $\boldsymbol{Q}_m \in \mathbb{C}^{2N_r\times(2N_r-N_t)}$ as the matrix containing the $2N_r - N_t$ right singular vector of $[\boldsymbol{G}_m^H, -\boldsymbol{G}_{m'}^H]$ corresponding to its zero singular values. The combiner of the NOMA users in the m-th group can be calculated as

$$[\boldsymbol{v}_m^H, \boldsymbol{v}_{m'}^H]^H = \boldsymbol{Q}_m \boldsymbol{x}_m, \tag{7}$$

where $\boldsymbol{x}_m$ is a random vector that generates a vector orthogonal to $[\boldsymbol{G}_m^H, -\boldsymbol{G}_{m'}^H]$ [13]. Aided by the combiner, (6) is reduced to $[\boldsymbol{G}_1^H \boldsymbol{v}_1, \boldsymbol{G}_2^H \boldsymbol{v}_2, ..., \boldsymbol{G}_{\frac{K_2}{2}}^H \boldsymbol{v}_{\frac{K_2}{2}}]^H \boldsymbol{w}_i = \boldsymbol{0}_{\frac{K_2}{2}\times 1}$, and the DoF of the precoder is less constrained compared to that in (6). On the other hand, for suppressing the MUI among the NOMA groups, we impose condition such that

$$[\boldsymbol{G}_1^H \boldsymbol{v}_1, ..., \boldsymbol{G}_{m-1}^H \boldsymbol{v}_{m-1}, \boldsymbol{G}_{m+1}^H \boldsymbol{v}_{m+1}, ..., \boldsymbol{G}_{K_2/2}^H \boldsymbol{v}_{K_2/2}]^H \boldsymbol{w}_m = \boldsymbol{0}_{(K_2/2-1)\times N_t}, \forall m. \tag{8}$$

(6) and (8) denote that the interference from the CI users and interference from other NOMA groups has been mitigated. Our next step is to allocate necessary power for each NOMA group for achieving their SINR requirements. Define τ_m as the equivalent power allocated for the m-th NOMA group, i.e., $\tau_m = [\boldsymbol{G}_m^H \boldsymbol{v}_m]^H \boldsymbol{w}_m \triangleq [\boldsymbol{G}_{m'}^H \boldsymbol{v}_{m'}]^H \boldsymbol{w}_m$. The SINR of the weak user for decoding its own signal is simplified as

$$\Gamma_{m'} = \frac{|\tau_m \rho_{m'} b_{m'}|^2/d_{m'}^{\lambda}}{|\tau_m \rho_m b_m|^2/d_{m'}^{\lambda} + ||\boldsymbol{v}_{m'}^H||^2\sigma^2}, \tag{9}$$

while the SINR of the strong user for decoding the weak user's signal is given as

$$\Gamma_{m'm} = \frac{|\tau_m \rho_{m'} b_{m'}|^2/d_m^\lambda}{|\tau_m \rho_m b_m|^2/d_m^\lambda + ||\boldsymbol{v}_m^H||^2\sigma^2}. \tag{10}$$

Since τ_m denotes the power allocated for the m-th NOMA group, the values of $\Gamma_{m'}$ and $\Gamma_{m'm}$ are determined by the path loss, i.e., d_m^λ and $d_{m'}^\lambda$. Since the strong user has a smaller value of path loss (the ambiguity caused by the small-scale fading has been eliminated by τ_m), $\Gamma_{m'm} \geq \Gamma_{m'}$ always holds, which means constraint (C3) is naturally guaranteed. Thus, the post-SIC SINR of the strong user is simplified as

$$\Gamma_m = \frac{|\tau_m \rho_m b_m|^2/d_m^\lambda}{||\boldsymbol{v}_m^H||^2\sigma^2}, \tag{11}$$

where we are able to link the power-related variable τ_m to the subscribed SINR $\bar{\Gamma}_m$ in (C2), written as

$$\tau_m = \max\{(\frac{\bar{\Gamma}_m||\boldsymbol{v}_m^H||^2\sigma^2 d_m^\lambda}{\rho_m^2})^{\frac{1}{2}}, (\frac{\bar{\Gamma}_m||\boldsymbol{v}_{m'}^H||^2\sigma^2 d_{m'}^\lambda}{\rho_{m'}^2 - \rho_m^2})^{\frac{1}{2}}\}, \forall m, \tag{12}$$

Finally, based on (8) and (12), we reach a generalized form for the precoder matrix $\boldsymbol{W}$, written as

$$[\boldsymbol{G}_1^H \boldsymbol{v}_1, ..., \boldsymbol{G}_{\frac{K_2}{2}}^H \boldsymbol{v}_{\frac{K_2}{2}}]^H \boldsymbol{W} = [\boldsymbol{0}_{\frac{K_2}{2}\times K_1}, \mathrm{diag}(\tau_1, ..., \tau_{\frac{K_2}{2}})]. \tag{13}$$

Now, P1 can be equivalently reformulated as

$$\begin{aligned} &P2 : \underset{\boldsymbol{W}, \gamma_i, \forall i \in \mathbb{K}_1}{\arg\max} \ \min \gamma_i, \\ &(\mathrm{C4}) : ||\boldsymbol{W}\boldsymbol{s}||^2 \leq N_t p_{\max}, \ (\mathrm{C5}) : (5), \ (\mathrm{C6}) : (13), \end{aligned} \tag{14}$$

where (C5) guarantees the MUI as a beneficial element for all the CI users. (C6) mitigates the MUI for all the NOMA-based users and meanwhile ensures their subscribed SINR requirements. Since now P2 is a second order cone programming (SOCP), it can be readily solved by commercial solvers. The CNOMA precoder directly positions the received signal of each user into correct regions of signal demodulation, without involving SDP. Hence, the subsequent semi-definite matrix decomposition and receiver-side phase equalization are not required, unlike other optimization-based precoders [2] [3]. The design of CNOMA precoder is outlined in Algorithm I.

Algorithm 1 The CNOMA Precoder Design

Input: Channel state information (CSI), power budget $p_{\max}$, and SINR threshold $\bar{\Gamma}_m$ for each NOMA group.
1: Design combiners for the NOMA-based users according to (7), and use equal-gain or SVD based combiner for the CI users.
2: Calculate power-related variable τ_m based on (12).
3: Solve the precoding optimization P2.
Output: Optimal precoding matrix $\boldsymbol{W}^*$.

The optimization of CNOMA precoder is subject to $2K_1$ linear constraints in (C5) with size 1, $K_2 N_t/2$ linear constraint in (C6) with size 1, and 1 second order cone constraint in (C4) with size N_t^2. Hence, it can be solved directly, without requiring intermediate parameters iteration or multi-layer optimization for convergence. Its complexity is computed as $\mathrm{C} = \sqrt{2K_1 + N_t K_2/2 + 2} \cdot \big(\epsilon(2K_1 + N_t K_2/2) + \epsilon^2(2K_1 + N_t K_2/2) + \epsilon N_t^4) + \epsilon^3\big)$, where $\epsilon = \mathcal{O}(N_t^2)$.

Remark 1: When the number of the users being served is larger than that of the transmit-antennas, it may not be able to provide constructive MUI for all the users. One needs to strike the fundamental tradeoff between utilizing MUI and improving multiuser access capability. That is, when accommodating more NOMA-based users, the ability of utilizing interference is reduced, and vice versa. □

Remark 2: The CNOMA precoder enables a tunable performance between the MUI utilization and MU access capability. It reduces to the CI precoder when $K_2 = 0$, while becomes to the NOMA precoder when $K_1 = 0$. Also, a larger value of τ_m, denotes that more power is assigned to the NOMA users. Though the interference from the NOMA users contributes constructively to the CI users, a larger value of τ_m still inevitably degrades the performance of the CI users. Also, with a low level of noise, a smaller value of τ_m is needed to guarantee the NOMA users' SINR requirements, and thus more power can be allocated for the CI users. □

Remark 3: Practical systems suffer imperfect CSI and SIC. CSI and SIC error can be formulated as bounded or unbounded model, where deterministic or probabilistic robust optimization can be formulated, respectively. □

IV. Simulation

Monte-Carlo simulation results are demonstrated in this section. Quadrature phase shift keying (QPSK) is adopted for modulation, and we assume each block consists of 40 symbols. The BS is equipped with $N_t = 10$ transmit-antennas while each user is equipped with $N_r = 6$ receive-antennas. Without loss of generality, the BS serves $K_1 = 3$ CI-based users and $K_2 = 14$ NOMA-based users. Rayleigh block fading is considered for small-scale channel model [13] [15]. The path loss exponent λ is set to as 2. The distance to the BS is set to as 20m for strong users and 30m for weak users, while the distance between the BS and CI-based users is randomly distributed in 20∼30m [13]. The power parameters of each NOMA group are set to as $\rho_m^2 = 0.25$ for strong user and $\rho_{m'}^2 = 0.75$ for weak user. For the purpose of simplicity, we set the power-related variable $\tau_m = 0.01$, $\forall m$, which corresponds to around -20 dB SINR threshold of the NOMA users. The following closely-related precoders are selected as benchmarks: 1) CI precoder [15], 2) MMSE precoder [1], and 3) NOMA MIMO precoder [13].

In Fig. 2(a), the impact of transmission power on the symbol error rate (SER) is presented, where the CNOMA precoder outperforms the CI/MMSE precoders at all transmission power. By utilizing MUI as a constructive element, the CNOMA precoder endorses the lowest SER for the CI-based users, achieving at most 0.6 Watt performance gain over the NOMA precoder. On the other hand, with a power threshold τ_m, the SER of the NOMA-based users is maintained at a reasonable level. In Fig. 2(b), the impact of power on the throughput is shown. The CNOMA obtains 200% throughput enhancement over the CI/MMSE and 12% enhancement over the NOMA precoder. With a small budget, e.g., $p_{\max} \leq 0.4$

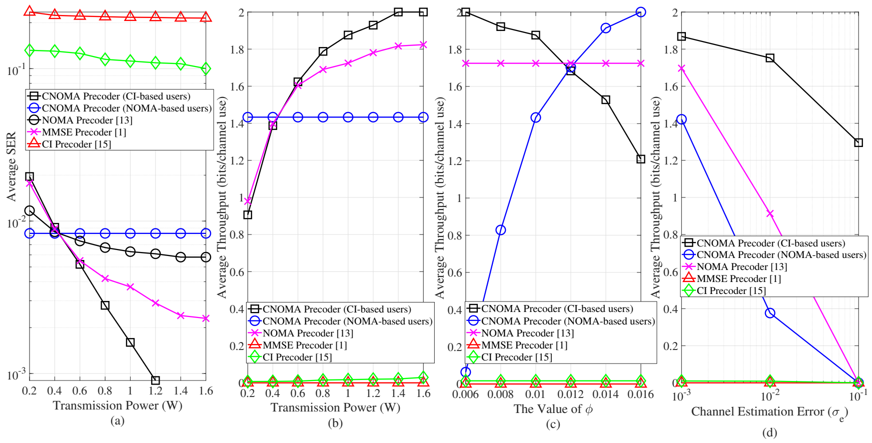

Fig. 2. The impact of power budget p_{max}, power threshold τ_m, and channel estimation quality on SER and throughput performance.

Watt, a large portion of power has to be assigned to the NOMA users for maintaining their SINR. When power increases, the CI users begin to outperform the NOMA users, and finally saturates at 2 bits/channel use, the highest throughput under QPSK. In Fig. 2(c), the impact of power threshold τ_m on the throughput is demonstrated. It is seen that the CNOMA provides a tunable performance for the CI- and NOMA-based users, by adjusting the value of τ_m. For example, a higher value of τ_m, i.e., $\tau_m \geq 0.012$, denotes that more power is allocated for the NOMA users, enabling the NOMA users to outperform the CI users. In Fig. 2(d), the impact of channel estimation quality on the throughput is presented. Define $\boldsymbol{H}_e$ as the channel estimation error matrix, and its element follows Gaussian distribution, i.e., $\mathcal{CN}\{0, \sigma_e^2\}$. By the CNOMA design, high throughput is always maintained for the CI users, providing high robustness against channel uncertainty. Also, the performance of the NOMA users is generally sensitive to the estimation quality. It is because imperfect CSI impairs the SIC operation, increasing the level of the residual interference. Hence, one may accordingly improve the SINR requirement for the NOMA users given a poor estimation quality.

V. CONCLUSIONS

In this paper, a novel constructive NOMA mechanism has been demonstrated. Aided by the synthetic successive coding and hybrid MU access, it ensures that the composite MUI acts as a beneficial element for the users having high throughput requirement, while suppressing interference for the NOMA-based users to guarantee their subscribed SINR performance. Then, a novel CNOMA precoder has been proposed, which presents a tunable performance between the exploitation of constructive MUI and enhancement of the MU access capability. Finally, simulation results have confirmed the superiority of the proposed CNOMA precoder over the classic CI, MMSE, and NOMA precoders, in terms of SER and throughput. Constructive NOMA is still broadly open, and future directions include but are not limited to closed-form precoder, imperfect CSI/SIC aware design, and CI-based rate-splitting design.

REFERENCES

[1] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[2] M. F. Hanif, L.-N. Tran, A. Tolli, and M. Juntti, "Computationally efficient robust beamforming for SINR balancing in multicell downlink with applications to large antenna array systems," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 1908–1920, Jun. 2014.

[3] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.

[4] Z. Wei *et al.*, "Fundamentals of physical layer anonymous communications: sender detection and anonymous precoding," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 64–79, Jan. 2022.

[5] Z. Ding, R. Schober, and H. V. Poor, "Hybrid NOMA offloading in multi-user MEC networks," *IEEE Trans. Wireless Comms.*, early access.

[6] Z. Wei, X. Zhu, S. Sun, J. Wang, and L. Hanzo, "Energy-efficient full-duplex cooperative non-orthogonal multiple access, *IEEE Trans. Veh. Technol.,* vol. 67, no. 10, pp. 10123-10128, Oct. 2018.

[7] M. F. Hanif *et al.*, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Trans. Sig. Process.*, vol. 64, no. 1, pp. 76-88, Jan. 2016.

[8] B. Clerckx *et al.*, "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, Feb, 2021.

[9] Y. Chang and K. Fukawa, "Non-orthogonal multiple access with phase rotation employing joint MUD and SIC," *Proc. VTC'18,* Porto, Portugal.

[10] A. Kabiri, M. J. Emadi, and M. N. Khormuji, "Optimal design of semi-orthogonal multiple-access massive MIMO systems, *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2230-2233, Oct. 2017.

[11] Q. Sun *et al.*, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.,* vol. 4, no. 4, pp. 405-408, Aug. 2015.

[12] 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.

[13] Z. Ding, R. Schober, and H. V. Poor, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Comms.*, vol. 15, no. 6, pp. 4438-4453, Jun. 2016.

[14] C. Masouros *et al.*, "Known interference in the cellular downlink: a limiting factor or a potential source of green signal power?" *IEEE Commun. Mag.*, vol. 51, no. 10, pp. 162–171, Oct. 2013.

[15] Z. Wei *et al.*, "Multi-cell interference exploitation: enhancing the power efficiency in cell coordination" *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 547-562, Jan. 2020.

[16] S. M. R. Islam *et al.*, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: potentials and challenges," *IEEE Commun. Sur. Tut.,* vol, 19, issue: 2, pp. 721-742, Oct. 2017.

[17] A. Jee, K. Agrawal, and S. Prakriya, "A coordinated direct AF/DF relay-aided NOMA framework for low outage," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 1559-1579, Mar. 2022.