Utilizing Imperfect Resolution of Near-Field Beamforming: A Hybrid-NOMA Perspective

Zhiguo Ding, *Fellow, IEEE* and H. Vincent Poor, *Life Fellow, IEEE*

Abstract—This letter studies how the imperfect resolution of near-field beamforming, the key feature of near-field communications, can be used to improve the throughput and connectivity of wireless networks. In particular, a hybrid non-orthogonal multiple access (NOMA) transmission strategy is developed to use preconfigured near-field beams for serving additional users. An energy consumption minimization problem is first formulated and then solved by using different successive interference cancellation strategies. Both analytical and simulation results are presented to illustrate the impact of the resolution of near-field beamforming on the design of hybrid NOMA transmission.

Index Terms—Near-field communication, non-orthogonal multiple access (NOMA), resolution of near-field beamforming.

I. Introduction

Near-field communication has recently received considerable attention because of its innovative way of utilizing spatial degrees of freedom [1]–[4]. Different from conventional far-field beamforming which relies on the approximated planar channel model, near-field beamforming is built on a more accurate spherical model. The recent studies in [5], [6] showed that the use of the spherical model enables beamfocusing, i.e., users with different locations can be distinguished in the so-called distance-angle domain, whereas conventional far-field beamforming relies on the angle domain only. When the users' distances to the base station are proportional to the Rayleigh distance, the resolution of near-field beamforming, i.e., whether users can be distinguished in the distance-angle domain, has been shown to be poor [7]. However, if the users are extremely close to the base station, indeed the resolution of near-field beamforming is almost perfect, as shown in Fig. 1(a) [8]. When there are multiple users that are very close to the base station, Fig. 1(b) shows that near-field beamforming can still effectively suppress the interference among these users. However, Fig. 1(b) also reveals an important observation that the magnitude of these users' near-field beams can be accumulated at a far location, which presents an opportunity to use these preconfigured near-field beams for supporting additional users.

Motivated by this observation, this letter focuses on the scenario with multiple near-field legacy users, with the aim of studying how the near-field beams preconfigured to the legacy users can be used to serve additional users. The additional users can be served at the same time as the legacy users, as in those existing strategies based on pure non-orthogonal multiple access (NOMA) in [2], [7]. Different from these existing works, a hybrid NOMA strategy is considered in this letter, where the existing pure NOMA strategies and conventional orthogonal multiple access (OMA) belong to this general framework. We note that hybrid NOMA has been applied to the uplink scenario only [9], whereas the downlink

Z. Ding is with Khalifa University, Abu Dhabi, UAE, and University of Manchester, Manchester, M1 9BB, UK. H. V. Poor is with Princeton University, Princeton, NJ 08544, USA.
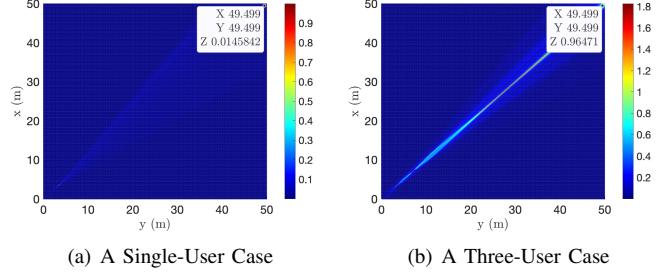

(a) A Single-User Case   (b) A Three-User Case

Fig. 1. Illustration of the beamfocusing resolution (Δ), with a 513-antenna uniform linear array, where the carrier frequency is $f_c = 28$ GHz, the antenna spacing is $d = \frac{\lambda}{2}$ and λ denotes the wavelength. For Fig. 1(a), there is a single user located at $\left(5, \frac{\pi}{4}\right)$, and for Fig. 1(b), the users' locations are $\left(5, \frac{\pi}{4}\right)$, $\left(10, \frac{\pi}{4}\right)$, and $\left(40, \frac{\pi}{4}\right)$, respectively, where polar coordinates are used.

scenario is the focus of this letter. An energy consumption minimization problem is first formulated, and then solved based on different successive interference cancellation (SIC) strategies. Both analytical and numerical results are obtained for insightful understandings about hybrid NOMA assisted near-field communications. In particular, hybrid NOMA is shown to be preferable to pure NOMA in terms of energy consumption. In addition, when the resolution of near-field beamforming is perfect, it is preferable to use all the preconfigured beams to support additional users. However, if the resolution of near-field beamforming is poor, it is crucial to carry out beam selection for energy minimization.

II. System Model

Consider a downlink legacy network, where there exist M near-field users, which are denoted by U_m, $1 \leq m \leq M$, and served by a base station equipped with an N-antenna uniform linear array (ULA). The aim of the letter is to use the beams preconfigured to the legacy users to serve an additional user, denoted by U_0. Each user is assumed to have a single antenna. Consider that the ULA is placed at the center of a 2-dimensional plane. Denote the locations of U_m, the center of the ULA, and the n-th element of the ULA by $\boldsymbol{\psi}_m^U$, $\boldsymbol{\psi}_0$, and $\boldsymbol{\psi}_n$, respectively, where Cartesian coordinates are used. It is assumed that the legacy users are very close to the base station in order to facilitate the implementation of beamfocusing, whereas U_0 is located far away from the base station.

A. OMA Transmission

To avoid the interference between the two types of users, a straightforward OMA approach can be adopted by combining space division multiple access (SDMA) with time-division multiple access (TDMA). In particular, during the first M time slots, the base station broadcasts $\sum_{m=1}^{M} \mathbf{w}_m s_m$, and the legacy user U_m receives the following observation:

$$y_m = \mathbf{h}_m^H \sum_{m=1}^{M} \sqrt{P}\mathbf{w}_m s_m + n_m, \qquad (1)$$

where s_m, $\mathbf{w}_m$, P denote U_m's signal, beamforming vector and transmit power, respectively, n_m denotes the white noise with its power denoted by σ^2, U_m's spherical channel vector is denoted by $\mathbf{h}_m = \sqrt{N}\alpha_m \mathbf{b}\left(\boldsymbol{\psi}_m^{\mathrm{U}}\right)$, $\mathbf{b}\left(\boldsymbol{\psi}\right) = \frac{1}{\sqrt{N}}\left[e^{-j\frac{2\pi}{\lambda}|\boldsymbol{\psi}-\boldsymbol{\psi}_1|} \quad \cdots \quad e^{-j\frac{2\pi}{\lambda}|\boldsymbol{\psi}-\boldsymbol{\psi}_N|}\right]^T$, $\alpha_m = \frac{\lambda}{4\pi|\boldsymbol{\psi}_m^{\mathrm{U}}-\boldsymbol{\psi}_0|}$, and λ denotes the wavelength [1], [3], [4]. Therefore, U_m's data rate in OMA can be expressed as: $R_m^{\mathrm{OMA}} = \log\left(1 + \frac{P|\mathbf{h}_m^H \mathbf{w}_m|^2}{P\sum_{i=1,i\neq m}^M |\mathbf{h}_m^H \mathbf{w}_i|^2 + \sigma^2}\right)$.

During the $(M+1)$-th time slot, only U_0 is served with the following data rate:

$$R_0^{\mathrm{OMA}}\left(P_0^{\mathrm{OMA}}\right) = \log\left(1 + \frac{P_0^{\mathrm{OMA}}}{\sigma^2}|\mathbf{h}_0^H \mathbf{w}_0|^2\right), \qquad (2)$$

where $\mathbf{h}_0$, P_0^{OMA} and $\mathbf{w}_0$ denote U_0's channel vector, transmit power, and beamforming vector, respectively.

A conventional design of $\mathbf{w}_m$ is to apply the zero-forcing (ZF) approach, which means that $\mathbf{W} = \mathbf{H}\left(\mathbf{H}^H \mathbf{H}\right)^{-1}\mathbf{D}$, where $\mathbf{W} = \begin{bmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_M\end{bmatrix}$, $\mathbf{H} = \begin{bmatrix}\mathbf{h}_1 & \cdots & \mathbf{h}_M\end{bmatrix}$ and $\mathbf{D} = \mathrm{diag}\left(\mathrm{diag}\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right)\right)^{-\frac{1}{2}}$. If the resolution of near-field beamforming is perfect, i.e., $\Delta \triangleq \left|\mathbf{b}\left(\boldsymbol{\psi}_m^{\mathrm{U}}\right)^H \mathbf{b}\left(\boldsymbol{\psi}_n^{\mathrm{U}}\right)\right|^2 \to 0$ for $m \neq n$, the beamfocusing based beamforming can also be used, i.e., $\mathbf{w}_m = \mathbf{b}(\phi_m^{\mathrm{U}})$. For U_0, $\mathbf{w}_0$ can be simply be $\mathbf{w}_0 = \mathbf{b}(\phi_0^{\mathrm{U}})$. We note that U_0 might be a far-field user, where $\mathbf{w}_0 = \mathbf{b}(\phi_0^{\mathrm{U}})$ is still applicable since the far-field planar channel model is an approximation of the more accurate spherical model.

B. Hybrid NOMA Transmission

The key idea of hybrid NOMA is to use those beams preconfigured to the legacy users, $\mathbf{w}_m$, as bandwidth resources to serve U_0. In particular, during the first M time slots, the base station sends the following signal

$$x^{\mathrm{NOMA}} = \sum_{m=1}^M \mathbf{w}_m(\sqrt{P}s_m + f_m\sqrt{P_m}s_0), \qquad (3)$$

where P_m is U_0's transmit power allocated on $\mathbf{w}_m$, and f_m is the corresponding complex-valued coefficient whose magnitude is one. Because U_0 suffers large path losses and hence has poor channel conditions, U_0 is to directly decode its own signal. However, the legacy users, U_m, can have two decoding approaches, as illustrated in the following.

1) Approach I: Each legacy user decodes its own signal directly, which means that its data rate is given by

$$R_{1,m}^{\mathrm{I}} = \log\left(1 + \frac{P|\mathbf{h}_m^H \mathbf{w}_m|^2}{I_m + \sigma^2}\right), \qquad (4)$$

where $I_m = P\sum_{i=1,i\neq m}^M |\mathbf{h}_m^H \mathbf{w}_i|^2 + \left|\mathbf{h}_m^H \sum_{k=1}^M \mathbf{w}_k f_k \sqrt{P_k}\right|^2$. To avoid any performance degradation at the legacy users, U_0's transmit power needs to be capped; such that $R_{1,m}^{\mathrm{I}} \geq R_t$, $1 \leq m \leq M$, where R_t denotes the legacy user's target data rate. With this approach, U_0's data rate during the first M time slots is given by

$$R_{1,0}^{\mathrm{I}} = \log\left(1 + \frac{\left|\mathbf{h}_0^H \sum_{k=1}^M \mathbf{w}_k f_k \sqrt{P_k}\right|^2}{P\sum_{i=1}^M |\mathbf{h}_0^H \mathbf{w}_i|^2 + \sigma^2}\right). \qquad (5)$$

2) Approach II: Due to their strong channel conditions, the legacy users can first decode U_0's signal, before decoding their own signals. In particular, the data rate for U_m to decode U_0's signal during the first M time slots is capped as follows:

$$R_{1,m}^{\mathrm{II}} = \log\left(1 + \frac{\left|\mathbf{h}_m^H \sum_{k=1}^M \mathbf{w}_k f_k \sqrt{P_k}\right|^2}{P\sum_{i=1}^M |\mathbf{h}_m^H \mathbf{w}_i|^2 + \sigma^2}\right), \qquad (6)$$

for $1 \leq m \leq M$. After U_0's signal is decoded and subtracted, U_m can decode its own signal with the same data rate as the one in OMA, i.e., R_m^{OMA}. Because of the constraints shown in (6), U_0's data rate during the first M time slots is given by

$$R_{1,0}^{\mathrm{II}} = \min\left\{R_{1,0}^{\mathrm{I}}, R_{1,m}^{\mathrm{II}}, 1 \leq m \leq M\right\}. \qquad (7)$$

III. Resource Allocation for Hybrid-NOMA

With hybrid-NOMA, U_0 has access to the first M time slots, which is useful for U_0 to increase its data rate, but potentially increases its energy consumption. Therefore, this letter focuses on the following energy minimization problem:

$$\min_{P_m \geq 0, f_m} \quad MT\sum_{m=1}^M P_m + TP_0 \qquad (\text{P1a})$$

$$s.t. \quad MTR_{1,0}^k + TR_0^{\mathrm{OMA}}\left(P_0\right) \geq TR_0, \qquad (\text{P1b})$$

$$\left(R_{1,m}^{\mathrm{I}} - R_t\right)\delta \geq 0, 1 \leq m \leq M \qquad (\text{P1c})$$

for $k \in \{\mathrm{I}, \mathrm{II}\}$, where $\delta = 1$ for Approach I, $\delta = 0$ for Approach II, T denotes the duration of each time slot, P_0 denotes U_0's transmit power in the $(M+1)$-th time slot, and R_0 denotes U_0's target data rate. Depending on which of the two approaches is adopted, problem P1 can be solved differently, as shown in the following.

Remark 1: By setting $P_m = 0$, $1 \leq m \leq M$, hybrid NOMA becomes pure OMA. By setting $P_0 = 0$, hybrid NOMA becomes pure NOMA. Therefore, hybrid NOMA is a general framework with pure NOMA and OMA as its special cases.

A. Approach I

Note that for beamfocusing, $\mathbf{w}_m^H \mathbf{w}_n \approx 0$, for $m \neq n$, if the resolution of near-field beamforming is perfect, and $\mathbf{w}_m^H \mathbf{w}_n = 0$ if ZF beamforming is used. Therefore, $R_{1,m}^{\mathrm{I}}$ can be simplified as follows:

$$R_{1,m}^{\mathrm{I}} \approx \log\left(1 + \frac{P|\mathbf{h}_m^H \mathbf{w}_m|^2}{\left|\mathbf{h}_m^H \mathbf{w}_m f_m \sqrt{P_m}\right|^2 + \sigma^2}\right) \qquad (8)$$

$$= \log\left(1 + \frac{Ph_m}{h_m P_m + \sigma^2}\right),$$

where $h_m = \left|\mathbf{h}_m^H \mathbf{w}_m\right|^2$. Because $R_{1,m}^{\mathrm{I}}$ is not a function of f_m, f_m can be set as follows: $f_m = \frac{\mathbf{h}_0^H \mathbf{w}_m}{|\mathbf{h}_0^H \mathbf{w}_m|}$, which means that U_0's data rate can be expressed as follows:

$$R_{1,0}^{\mathrm{I}} = \log\left(1 + a_0\left|\sum_{k=1}^M g_k \sqrt{P_k}\right|^2\right), \qquad (9)$$

where $g_m = |\mathbf{h}_0^H \mathbf{w}_m|$, $0 \leq m \leq M$, and $a_m = \frac{1}{P\sum_{i=1}^M |\mathbf{h}_m^H \mathbf{w}_i|^2 + \sigma^2}$, $0 \leq m \leq M$.

By using the simplified expression for the data rates, problem P1 can be simplified as follows:

$$\min_{P_m \geq 0} M \sum_{m=1}^{M} P_m + P_0 \tag{P2a}$$

$$s.t. \ M \log \left(1 + a_0 \left| \sum_{k=1}^{M} g_k \sqrt{P_k} \right|^2 \right) + \log \left(1 + b_0 P_0 \right) \geq R_0 \tag{P2b}$$

$$P_m \leq \frac{P}{e^{R_t} - 1} - \frac{\sigma^2}{h_m}, 1 \leq m \leq M, \tag{P2c}$$

where $b_0 = \frac{g_0}{\sigma^2}$.

The constraint function in (P2b) is concave, as illustrated in the following. The left-hand side of (P2b) can be expressed as follows: $f(h(\mathbf{p}), P_0)$, where $\mathbf{p} = \begin{bmatrix} g_1^2 P_1 & \cdots & g_M^2 P_M \end{bmatrix}$, $f(x, P_0) = M \log(1 + a_0 x) + \log(1 + b_0 P_0)$, and $h(\mathbf{x}) = \left(\sum_{k=1}^{M} x_k^{\frac{1}{2}} \right)^2$. Note that $h(\mathbf{x})$ is simply the ℓ_p-norm of $\mathbf{x}$ with $p = \frac{1}{2}$. Recall that for $\mathbf{x} \geq 0$, its ℓ_p-norm is a concave function, if $p < 1$ [10]. Therefore, $f(h(\mathbf{p}), P_0)$ is the composition of two concave functions, and hence the constraint in (P2b) is concave. Therefore, problem P2 is a concave optimization problem, since both the objective function in (P2a) and the constraint in (P2c) are affine functions.

Remark 2: Note that P_m is non-negative, and hence the right-hand side of the constrain in (P2c) needs to be non-negative, i.e.,

$$\frac{P}{e^{R_t} - 1} \geq \frac{\sigma^2}{h_m} \implies \log \left(1 + \frac{P h_m}{\sigma^2} \right) \geq R_t. \tag{10}$$

which means that not all the beams, $\mathbf{w}_m$, are useful to U_0. In particular, beam selection needs to be carried out to ensure that only the qualified beams, i.e., $h_m \geq \sigma^2 \frac{e^{R_t} - 1}{P}$, are used to serve U_0.

B. Approach II

Again by applying the orthogonality between the preconfigured beams, U_m's data rate expression, $1 \leq m \leq M$, can be simplified as follows:

$$R_{1,m}^{II} \approx \log \left(1 + a_m \left| \mathbf{h}_m^H \mathbf{w}_m f_m \sqrt{P_m} \right|^2 \right) \tag{11}$$

$$= \log \left(1 + a_m h_m P_m \right).$$

Similar to Approach I, $R_{1,m}^{II}$ is also not a function of f_m, and hence the choice of $f_m = \frac{\mathbf{h}_0^H \mathbf{w}_m}{|\mathbf{h}_0^H \mathbf{w}_m|}$ is still applicable, which means that problem P1 can be simplified as follows:

$$\min_{P_m \geq 0} M \sum_{m=1}^{M} P_m + P_0 \tag{P3a}$$

$$s.t. \ M \min \left\{ \log \left(1 + a_0 \left| \sum_{k=1}^{M} g_k \sqrt{P_k} \right|^2 \right) \tag{P3b} \right.$$

$$\left. , \log \left(1 + a_m h_m P_m \right), 1 \leq m \leq M \right\}$$
$$+ \log \left(1 + b_0 P_0 \right) \geq R_0,$$

Problem P3 can be further recast as follows:

$$\min_{P_m \geq 0} M \sum_{m=1}^{M} P_m + P_0 \tag{P4a}$$

$$s.t. \ M \log \left(1 + a_m h_m P_m \right) + \log \left(1 + b_0 P_0 \right) \geq R_0, \tag{P4b}$$
$$1 \leq m \leq M,$$

$$M \log \left(1 + a_0 \left| \sum_{k=1}^{M} g_k \sqrt{P_k} \right|^2 \right) + \log \left(1 + b_0 P_0 \right) \geq R_0. \tag{P4c}$$

Similar to problem P2, problem P4 is also a concave optimization problem, and hence can be efficiently solved.

Remark 3: Similar to Approach I, beam selection could be also important to Approach II, as explained in the following. By inviting more beams to help U_0, U_0's effective channel gain, $\left| \sum_{k=1}^{M} g_k \sqrt{P_k} \right|^2$ in (P4c), is improved; however, the number of constraints in (P4b) is also increased, which can reduce the search space for the power allocation coefficients and hence cause performance degradation. One effective way for beam selection is to first order the beams according to h_m, and select the M_x best beams, $M_x \leq M$.

Remark 4: We note that the use of off-shelf optimization solvers can yield the optimal solution of problem (P4), but without insightful understandings to the properties of the solutions, which motivates the following special-case studies.

Lemma 1. *Assuming that the preconfigured beams are strictly orthogonal, the high-SNR approximations of the optimal choices of P_0 and P_m, $1 \leq m \leq M_x$, are given by*

$$P_1^* \approx \cdots \approx P_{M_x}^* \approx \frac{\lambda}{M_x} - \frac{1}{c}, \quad P_0^* \approx \lambda - \frac{1}{b_0}, \tag{12}$$

for $\frac{P}{\sigma^2} \to \infty$, if $R_0 \geq \frac{M_x b_0}{c}$, otherwise $P_1^* = \cdots = P_{M_x}^* = 0$, and $P_0^* = \frac{e^{R_0} - 1}{b_0}$, where the Lagrange multiplier is given by $\lambda^* = \left(\frac{M_x^M e^{R_0}}{c^M b_0} \right)^{\frac{1}{M+1}}$, and $c = \min \left\{ \frac{1}{P}, a_0 \left| \sum_{k=1}^{M_x} g_k \right|^2 \right\}$.

Proof. With the orthogonality assumption made in the lemma, $a_m h_m$ can be approximated at high SNR as follows:

$$a_m h_m = \frac{\left| \mathbf{h}_m^H \mathbf{w}_m \right|^2}{P \sum_{i=1}^{M} |\mathbf{h}_m^H \mathbf{w}_i|^2 + \sigma^2} \approx \frac{1}{P}, \tag{13}$$

where the orthogonality assumption ensures that $\mathbf{h}_m^H \mathbf{w}_i = 0$ for $m \neq i$. By using (13), the multiple constraints shown in (P4b) can be merged together, which means that $P_1 = \cdots = P_{M_x} \triangleq P_t$ at high SNR. Therefore, problem P4 can be recast as follows:

$$\min_{P_m \geq 0} M M_x P_t + P_0 \tag{P5a}$$

$$s.t. \ M \log \left(1 + \frac{P_t}{P} \right) + \log \left(1 + b_0 P_0 \right) \geq R_0, \tag{P5b}$$

$$M \log \left(1 + a_0 P_t \left| \sum_{k=1}^{M_x} g_k \right|^2 \right) + \log \left(1 + b_0 P_0 \right) \geq R_0,$$

which can be further simplified as follows:

$$\min_{P_m \geq 0} \quad MM_x P_t + P_0 \tag{P6a}$$

$$s.t. \quad M\log\left(1 + cP_t\right) + \log\left(1 + b_0 P_0\right) \geq R_0. \tag{P6b}$$

Problem P6 is a concave optimization problem, and its Lagrange can be expressed as follows:

$$L(P_0, P_t, \lambda) = MM_x P_t + P_0 \tag{14}$$
$$+ \lambda\left(R_0 - M\log\left(1 + cP_t\right) - \log\left(1 + b_0 P_0\right)\right),$$

where λ is a Lagrange multiplier [10]. By applying the Karush–Kuhn–Tucker (KKT) conditions, the lemma can be approved. $\square$

Remark 5: An immediate conclusion from Lemma 1 is that U_0's transmit powers during the first $M + 1$ time slots can be reduced by scheduling more beams, i.e., increasing M_x reduces P_m, as illustrated in the following. According Lemma 1, the optimal value for P_m can be approximated as follows:

$$P_m^* \approx \frac{1}{M_x}\left(\frac{M_x^M e^{R_0}}{c^M b_0}\right)^{\frac{1}{M+1}} - \frac{1}{c} \tag{15}$$
$$= \left(\frac{e^{R_0}}{M_x c^M b_0}\right)^{\frac{1}{M+1}} - \frac{1}{c} \approx \frac{1}{c}\left(\left(\frac{e^{R_0}}{M_x b_0}\right)^{\frac{1}{M+1}} - 1\right),$$

where the approximation is obtained by assuming large M. By increasing M_x, c is also increasing, which means that P_m^* is reduced. However, it is important to point out that the overall energy consumption may not be reduced by increasing M_x, as shown in the next section.

For the special case with $M = M_x = 1$, an exact expression of power allocation can be obtained without the high-SNR approximation, which leads to more insightful understandings.

Lemma 2. *For the special case with $M = M_x = 1$, the following conclusions can be made:*
- *The condition for hybrid NOMA to outperform pure OMA is $R_0 \geq \frac{b_0}{\beta}$;*
- *The pure NOMA solution, $P_0 = 0$, causes more energy consumption than hybrid NOMA.*

where $\beta = \min\{a_1 h_1, a_0 g_1^2\}$.

Proof. For the two-user case, problem P4 can be simplified as follows:

$$\min_{P_m \geq 0} \quad P_1 + P_0 \tag{P7a}$$

$$s.t. \quad \log\left(1 + a_1 h_1 P_1\right) + \log\left(1 + b_0 P_0\right) \geq R_0 \tag{P7b}$$
$$\log\left(1 + a_0 g_1^2 P_1\right) + \log\left(1 + b_0 P_0\right) \geq R_0.$$

By introducing β, the two constraints can be combined, and problem P7 can be further simplified as follows:

$$\min_{P_m \geq 0} \quad P_1 + P_0 \tag{P8a}$$

$$s.t. \quad \log\left(1 + \beta P_1\right) + \log\left(1 + b_0 P_0\right) \geq R_0. \tag{P8b}$$

It is straightforward to verify that problem P8 is a concave optimization problem, and the use of the KKT conditions leads to the following closed-form expressions of P_1 and P_0 [10]:

$$P_1^* = \sqrt{\frac{e^{R_0}}{\beta b_0}} - \frac{1}{\beta}, \quad P_0^* = \sqrt{\frac{e^{R_0}}{\beta b_0}} - \frac{1}{b_0}, \tag{16}$$
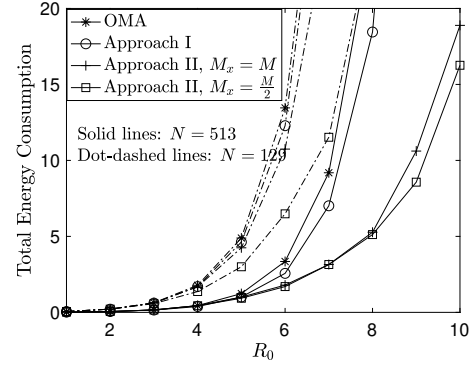

Fig. 2. Total energy consumption as a function of R_0 with randomly located users. $M = 40$, $P = 10$ dBm, $\sigma^2 = 70$ dBm, and $R_t = 4$ BPCU.

if $\sqrt{\frac{e^{R_0}}{\beta b_0}} - \frac{1}{\beta} \geq 0$, otherwise pure OMA is to be adopted. Establishing the condition for hybrid NOMA to outperform OMA is equivalent to identify the condition for $P_1^* > 0$. By using the closed-form expression of P_1^* shown in (16), the condition can be obtained, as shown in the lemma.

The second conclusion in the lemma can be proved by contradiction. In particular, the difference between pure NOMA and hybrid NOMA means that, if pure NOMA is adopted, the following condition must be satisfied:

$$P_0^* = \sqrt{\frac{e^{R_0}}{\beta b_0}} - \frac{1}{b_0} \leq 0, \tag{17}$$

which is equivalent to the following:

$$\frac{\beta}{b_0} \geq e^{R_0} \geq 1 \Longrightarrow \beta \geq b_0. \tag{18}$$

However, β is strictly larger than b_0, since

$$\beta = \min\left\{\frac{|\mathbf{h}_1^H \mathbf{w}_1|^2}{P|\mathbf{h}_1^H \mathbf{w}_1|^2 + \sigma^2}, \frac{|\mathbf{h}_0^H \mathbf{w}_1|^2}{P|\mathbf{h}_0^H \mathbf{w}_1|^2 + \sigma^2}\right\} \tag{19}$$
$$\leq \frac{|\mathbf{h}_0^H \mathbf{w}_1|^2}{P|\mathbf{h}_0^H \mathbf{w}_1|^2 + \sigma^2} < \frac{|\mathbf{h}_0^H \mathbf{w}_1|^2}{\sigma^2} \leq \frac{|\mathbf{h}_0^H \mathbf{w}_0|^2}{\sigma^2} = b_0,$$

which contradicts to the conclusion in (18). Therefore, the second part of the lemma is proved. $\square$

Remark 6: Lemma 2 indicates that pure NOMA suffers a performance loss to hybrid NOMA, which is consistent to the conclusion previously reported to the uplink scenario [9]. In addition, Lemma 2 shows that the use of hybrid NOMA is particularly useful for the case with large R_0.

IV. SIMULATION RESULTS

In this section, computer simulation results are presented to demonstrate the performance of hybrid NOMA transmission. For all the presented simulation results, $f_c = 28$ GHz and the antenna spacing is $d = \frac{\lambda}{2}$.

In Fig. 2, the performance of hybrid NOMA is studied by assuming that the users are randomly located. In particular, the legacy users are randomly located in a half-ring with its inner radius 5 m and its outer radius 10 m. U_0 is randomly located in a half-ring with its inner radius 150 m and its outer radius 200 m. Because the users are randomly located, the users' channel vectors may not be orthogonal, which motivates
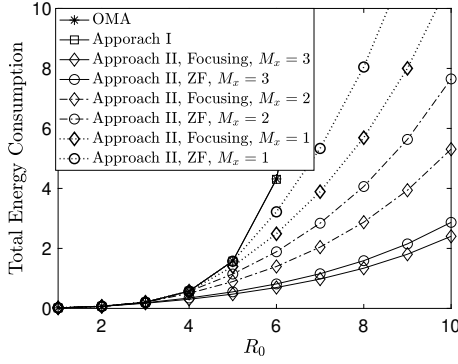
Fig. 3. The impact of M_x on the energy consumption of hybrid NOMA. $N = 513$, $M = 3$, $P = 10$ dBm, $\sigma^2 = 70$ dBm, and $R_t = 4$ BPCU.

Fig. 4. Impact of P and R_t on the total energy consumption. $N = 513$, $M = 3$, $M_x = 3$ $\sigma^2 = 70$ dBm. Beamfocusing is used for beamforming.

the use of ZF beamforming. As can be seen from the figure, the use of Approach II can reduce U_0's energy consumption significantly, compared to OMA, particularly for large R_0, whereas Approach I realizes a performance quite similar to OMA. By increasing the number of antennas of the ULA, the performance gap between hybrid NOMA and OMA can be further enlarged.

In Figs. 3 and 4, a deterministic scenario is focused on. In particular, U_m, $1 \le m \le 3$, are located at $\left(5, \frac{\pi}{4}\right)$, $\left(10, \frac{\pi}{4}\right)$, and $\left(40, \frac{\pi}{4}\right)$, respectively, U_0 is located at $\left(200, \frac{\pi}{4}\right)$, where polar coordinates are used. This scenario represents an important application of near-field communications, such as vehicular networks, where the users are expected to be lined up on a straight street. We note that for this deterministic scenario, it can be easily verified that the channel vectors of U_m, $1 \le m \le M$, are almost orthogonal to each other, which means that both beamfocusing and ZF can be used. Both the figures shows that hybrid NOMA can outperform OMA, particularly with large R_0, which confirms Lemma 2. In addition, Fig. 3 shows that Approach II outperforms Approach I with small P and large R_t. However, by increasing P and reducing R_t, Approach II can outperform Approach I, as shown in Fig. 4. Furthermore, Table I also demonstrates the accuracy of the high-SNR approximation developed in Lemma 1.

The key feature of the resolution of near-field beamforming causes an interesting observation that increasing M_x introduces a performance loss in Fig. 2, but yields a performance gain in Fig. 3, as explained in the following. For the random setup considered in Fig. 3, the users' channel vectors might be linear dependent, instead of orthogonal, to each other, which means that the use of ZF beamforming can cause the users' channel gains fluctuated. For example, assume that the users' channel vectors are orthogonal, e.g., $\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, which means that the users' power normalization coefficients are the same, i.e., diag$\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right) = \frac{1}{2}\begin{bmatrix} 1 & 1 \end{bmatrix}$. However, if their channel vectors are not orthogonal, e.g., $\mathbf{H} = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}$, the users' power normalization coefficients are no longer the same, i.e., diag$\left(\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right) = \begin{bmatrix} 1 & 2 \end{bmatrix}$. This fluctuation can make some users' beams not as useful as the others', which motivates the use of beam selection, as explained in Remark 3. However, for the case considered in Fig. 3, the resolution of near-field beamforming is perfect, i.e., the users' channel
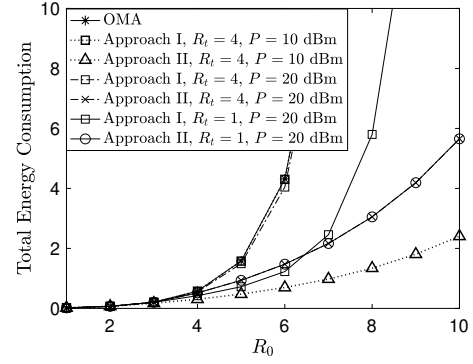
vectors are almost orthogonal to each other, and the fluctuation issue does not exist, which means that all the beams should be used, i.e., increasing M_x improves the system performance.

V. CONCLUSIONS

In this letter, a hybrid NOMA transmission strategy was developed to use the preconfigured near-field beams for serving additional users. An energy consumption minimization problem was first formulated and then solved by using different successive interference cancellation strategies. Both analytical and simulation results were presented to illustrate the impact of the resolution of near-field beamforming on the design of hybrid NOMA transmission.

TABLE I
POWER ALLOCATION FOR HYBRID NOMA (IN WATTS)

	P_1	P_2	P_3	P_4
Optimal ($\sigma^2 = -70$ dBm)	0.0684	0.1236	0.3875	0.6647
Approx. ($\sigma^2 = -70$ dBm)	0.2118	0.2118	0.2118	0.7338
Optimal ($\sigma^2 = -80$ dBm)	0.0430	0.0430	0.0430	0.1580
Approx. ($\sigma^2 = -80$ dBm)	0.0430	0.0430	0.0430	0.1579

REFERENCES

[1] J. Zhu, Z. Wan, L. Dai, M. Debbah, and H. V. Poor, "Electromagnetic information theory: Fundamentals, modeling, applications, and open problems," Available on-line at arXiv:2209.09562, 2022.

[2] Z. Ding, R. Schober, and H. V. Poor, "NOMA-based coexistence of near-field and far-field massive MIMO communications," *IEEE Wireless Commun. Lett.*, vol. 12, no. 8, pp. 1429–1433, 2023.

[3] H. Zhang, N. Shlezinger, F. Guidi, D. Dardari, M. F. Imani, and Y. C. Eldar, "Beam focusing for near-field multiuser MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7476–7490, Sept. 2022.

[4] X. Zhang, H. Zhang, and Y. C. Eldar, "Near-field sparse channel representation and estimation in 6G wireless communications," Available on-line at arXiv:2212.13527, 2022.

[5] J. Chen, F. Gao, M. Jian, and W. Yuan, "Hierarchical codebook design for near-field mmwave MIMO communications systems," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2023.

[6] Y. Liu, J. Xu, Z. Wang, X. Mu, and L. Hanzo, "Near-field communications: What will be different?" *IEEE Wireless Commun.*, Available on-line at arXiv:2303.04003, 2023.

[7] Z. Ding, "Resolution of near-field beamforming and its impact on NOMA," *IEEE Wireless Commun. Lett.*, Available on-line at arXiv:2308.08159, 2023.

[8] Z. Wu and L. Dai, "Multiple access for near-field communications: SDMA or LDMA?" Available on-line at arXiv:2208.06349, 2022.

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

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2003.