A Simple Design of IRS-NOMA Transmission

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

***Abstract*— This letter proposes a simple design of intelligent reflecting surface (IRS) assisted non-orthogonal multiple access (NOMA) downlink transmission. In particular, conventional spatial division multiple access (SDMA) is used first at the base station to generate orthogonal beams by using the spatial directions of the near users' channels. Then, IRS-assisted NOMA is used to ensure that additional cell-edge users can also be served on these beams by aligning the cell-edge users' effective channel vectors with the predetermined spatial directions. Both analytical and simulation results are provided to demonstrate the performance of the proposed IRS-NOMA scheme and also study the impact of hardware impairments on IRS-NOMA.**

***Index Terms*— Non-orthogonal multiple access (NOMA), intelligent reflecting surface (IRS), phase shifting design, diversity order.**

I. Introduction

NON-ORTHOGONAL multiple access (NOMA) has been recognized as a promising multiple access candidate for future mobile networks [1]. For example, NOMA has been continuously studied in the 3GPP framework. In 2015, a study for using NOMA in the downlink communication scenario, termed multi-user superposition transmission (MUST), was carried out for 3GPP Release 14 [2], which led to the inclusion of MUST in 3GPP Release 15, also termed Evolved Universal Terrestrial Radio Access (E-UTRA) [3]. In 2018, a further study for using NOMA in the uplink communication scenario was carried out for 3GPP Release 16, with more than 20 different NOMA uplink transmission schemes proposed by various industrial and academic groups [4]. The key idea of NOMA is to serve multiple users on each orthogonal bandwidth resource block. In scenarios with multiple-antenna nodes, orthogonal spatial directions can be viewed as a type of resource blocks. Conventional orthogonal multiple access (OMA), such as spatial division multiple access (SDMA), is to serve a single user on each spatial direction, whereas the use of NOMA can ensure that multiple users are served simultaneously on each spatial direction and hence improves spectral efficiency. However, it is important to point out that the use of NOMA is not always preferable [5]. For example, if users' channel vectors are orthogonal to each other, SDMA is more preferable than NOMA, whereas the situation, in which the directions of the users' channel vectors are the same, is the ideal case for the implementation of NOMA.

Manuscript received November 18, 2019; revised December 20, 2019 and February 3, 2020; accepted February 9, 2020. Date of publication February 17, 2020; date of current version May 8, 2020. Z. Ding's work was supported by the U.K. EPSRC under Grant EP/P009719/2, and in part by H2020-MSCA-RISE-2015 under Grant 690750. H. V. Poor's work was supported by the U.S. National Science Foundation under Grant CCF-1908308. The associate editor coordinating the review of this letter and approving it for publication was Q. Wu. *(Corresponding author: Zhiguo Ding.)*

Zhiguo Ding is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA, and also with the Department of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, U.K. (e-mail: zhiguo.ding@manchester.ac.uk).

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

Digital Object Identifier 10.1109/LCOMM.2020.2974196

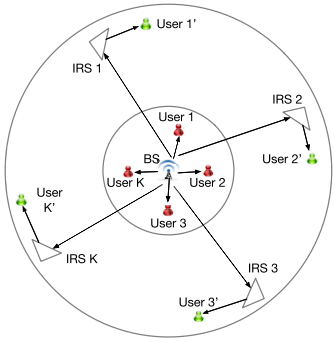


Fig. 1. A system diagram for IRS-NOMA with $2K$ users and K IRS's.

Therefore, an important question to broaden the applications of NOMA is whether the directions of users' channel vectors can be manipulated, i.e., aligning one user's channel with the others'. This is difficult in conventional wireless systems, since the users' channels are fixed and determined by propagation environments. Motivated by this difficulty, this letter is to propose a new type of NOMA transmission by employing the intelligent reflecting surface (IRS) which can be viewed as a low-cost antenna array consisting of a large number of reconfigurable reflecting elements [6]–[11]. By applying IRS, the direction of a user's channel vector can be effectively tuned, which facilitates the implementation of NOMA. In particular, the spectral efficiency and connectivity can be improved by IRS-NOMA since a single spatial direction can be used to serve multiple users, even if their original channels are not aligned. Due to the low-cost feature of IRS, e.g., finite-resolution phase shifters, a user's channel vector cannot be accurately aligned to a target direction. The impact of this hardware impairment on IRS-NOMA is investigated and the performance of the developed practical IRS-NOMA transmission scheme is characterized in this letter.

II. System Model

Consider a multi-user downlink scenario as shown in Fig. 1. There are two types of users, namely near users and cell-edge users, where it is assumed that there is no direct link between the base station and the cell-edge users. SDMA is used first, where the base station equipped with M antennas generates K beamforming vectors, denoted by $\mathbf{w}_k$, $1 \leq k \leq K$, to serve K ($M \geq K$) near users by applying zero forcing beamforming. To better illustrate the benefit of IRS-NOMA, we further assume that the K near users are scheduled because their channel vectors are orthogonal to each other, which means that the beamforming vectors, $\mathbf{w}_k$, are orthonormal vectors. After $\mathbf{w}_k$ is generated, IRS-NOMA is used to ensure that more users can be served on these predetermined beams than

conventional SDMA. For illustration purposes, we assume that on each beam, $\mathbf{w}_k$, one additional user,[1] denoted by user k', is served with the help of an IRS which is equipped with N reflecting elements, as shown in Fig. 1. In addition, we assume that only user k' can hear IRS k since the IRS's are deployed close to the cell-edge users.[2]

The base station broadcasts $\sum_{k=1}^{K} \mathbf{w}_k(\alpha_1 s_k + \alpha_2 s_{k'})$, where $s_{k'}$ denotes the signal to be sent to user k', s_k is the signal to be sent to user k, α_i denotes the power allocation coefficient, and $\alpha_1^2 + \alpha_2^2 = 1$. Because it is assumed that there is no direct link between the IRS's and the near users, the performance analysis for user k is exactly the same as those in conventional NOMA systems, and hence in this letter we focus on the performance at user k' only.

Therefore, the signal received by user k' is given by

$$y_{k'} = \mathbf{h}_{k'}^H \mathbf{\Theta}_k \mathbf{G}_k \sum_{k=1}^{K} \mathbf{w}_k(\alpha_1 s_k + \alpha_2 s_{k'}) + w_{k'}, \tag{1}$$

where $\mathbf{G}_k$ denotes the $N \times M$ complex Gaussian channel matrix from the base station to the IRS associated with user k', $\mathbf{h}_{k'}$ denotes the complex Gaussian channel vector from the IRS to user k', and $w_{k'}$ denotes the noise. $\mathbf{\Theta}_k$ is a diagonal matrix, and each of its main diagonal elements is denoted by $\beta_{k,i} e^{-j\theta_{k,i}}$, where $\theta_{k,i}$ denotes the reflection phase shift and $\beta_{k,i}$ denotes the amplitude reflection coefficient [6], [7]. We note that in this letter only small scale fading is considered, whereas large scale path loss is omitted, i.e., each element of $\mathbf{G}_k$ and $\mathbf{h}_{k'}$ is independent complex Gaussian distributed with zero mean and unit variance.

As in conventional NOMA, it is assumed that $\alpha_1 \leq \alpha_2$, and hence the signal-interference-plus-noise (SINR) for user k' to decode its message is given by

$$\text{SINR}_{k'} = \frac{|\boldsymbol{\theta}_k^H \mathbf{D}_{k'} \mathbf{h}_k|^2 \alpha_2^2}{|\boldsymbol{\theta}_k^H \mathbf{D}_{k'} \mathbf{h}_k|^2 \alpha_1^2 + \sum_{i=1, i\neq k}^{K} |\boldsymbol{\theta}_k^H \mathbf{D}_{k'} \mathbf{h}_i|^2 + \frac{1}{\rho}}, \tag{2}$$

where ρ denotes the transmit signal-to-noise ratio (SNR), $\mathbf{h}_i = \mathbf{G}_k \mathbf{w}_i$, $\boldsymbol{\theta}_k$ is an $N \times 1$ vector containing the elements on the main diagonal of $\mathbf{\Theta}_k^H$, and $\mathbf{D}_{k'}$ is a diagonal matrix with its diagonal elements obtained from $\mathbf{h}_{k'}^H$. We note that the SINR in (2) implies that a cell-edge user has the perfect knowledge of $\mathbf{h}_{k'}$, $\mathbf{G}_k$, and $\mathbf{w}_k$. The channel state information related to IRS, $\mathbf{G}_k$ and $\mathbf{h}_{k'}$, is assumed to be available via the channel estimation methods described in [12]–[14]. The information about the predetermined beamforming vectors, $\mathbf{w}_k$, can be sent to the user by its IRS via a reliable control channel. Furthermore, we note that the SINR expression is more complicated than that of conventional NOMA, due to the existence of the product of the complex Gaussian distributed random variables.

III. Designs of IRS-NOMA

As can be observed from (2), the application of IRS ensures that a cell-edge user's effective channel vector, $\mathbf{h}_{k'}^H \mathbf{\Theta}_k \mathbf{G}_k$, can be manipulated, which uses the key feature of IRS, i.e., making the environment controllable and programmable. In particular, the cell-edge user's effective channel vector is determined by the choice of $\mathbf{\Theta}_k$, and two designs of $\mathbf{\Theta}_k$ will be introduced in the following two subsections.

A. IRS-NOMA With Ideal Beamforming

Ideal designs of IRS-NOMA typically require beamforming with infinite resolution, i.e., the hardware circuit can support arbitrary choices for the phase shift and amplitude coefficient, $\theta_{k,i}$ and $\beta_{k,i}$. Take the zero-forcing design as an example. In order to suppress inter-pair interference, the use of zero-forcing beamforming implies that $\boldsymbol{\theta}_k$ should satisfy the following constraints:

$$\boldsymbol{\theta}_k^H \mathbf{D}_{k'} \mathbf{h}_i = 0, \tag{3}$$

for $i \neq k$. Given the fact that $\mathbf{w}_k$'s are orthogonal to each other, the constraint in (3) is equivalent to the one that a cell-edge user's effective channel vector $\mathbf{h}_{k'}^H \mathbf{\Theta}_k \mathbf{G}_k$ is aligned with $\mathbf{w}_k$. Denote $\mathbf{V}_k$ by an $N \times (N - K + 1)$ matrix collecting the basis vectors of the null space of $\left[\mathbf{D}_{k'}\mathbf{h}_1 \cdots \mathbf{D}_{k'}\mathbf{h}_{i-1}\ \mathbf{D}_{k'}\mathbf{h}_{i+1} \cdots \mathbf{D}_{k'}\mathbf{h}_K\right]$. Therefore, $\boldsymbol{\theta}_k$ can be obtained as $\mathbf{V}_k \mathbf{x}$, and the optimal $\mathbf{x}$ can be obtained as follows:

$$\max_{\mathbf{x}} \ \frac{|\mathbf{x}^H \mathbf{V}_k^H \mathbf{D}_{k'} \mathbf{h}_k|^2 \alpha_2^2}{|\mathbf{x}^H \mathbf{V}_k^H \mathbf{D}_{k'} \mathbf{h}_k|^2 \alpha_1^2 + \frac{1}{\rho}} \tag{4}$$

$$\text{s.t. } |\mathbf{x}|^2 \leq 1. \tag{5}$$

By using the fact that $\frac{\alpha_2^2\, y}{\alpha_1^2 y + \frac{1}{\rho}}$ is a mono-increasing function of y and also applying Cauchy-Schwarz inequality, the maximum of the SINR is achieved by $\boldsymbol{\theta}_k^* = \mathbf{V}_k \frac{\mathbf{V}_k^H \mathbf{D}_{k'} \mathbf{h}_k}{|\mathbf{V}_k^H \mathbf{D}_{k'} \mathbf{h}_k|}$. Evidently, such an ideal design requires that the number of possible choices for the phase shift and the amplitude, $\theta_{k,i}$ and $\beta_{k,i}$, is infinite. We note that the other ideal designs, e.g., directly maximizing the SINR in (2), lead to the same conclusion.

B. IRS-NOMA With Finite Resolution Beamforming

In practice, the choices for $\theta_{k,i}$ and $\beta_{k,i}$ cannot be arbitrary due to the hardware limitations. A straightforward design for IRS-NOMA with finite resolution beamforming is inspired by lens antenna arrays in millimeter-wave networks [15]. In particular, denote an $N \times N$ discrete Fourier transform (DFT) matrix by $\mathbf{F}_n$, and the optimal $\boldsymbol{\theta}_k$ to maximize the SINR in (2) can be found by an exhaustive search among the columns of $\mathbf{F}_N$, which can be realized by finite-resolution phase shiters.

An alternative low-cost implementation is to apply on-off control to IRS-NOMA, i.e., each diagonal element of $\mathbf{\Theta}$ is either 0 (off) or 1 (on). Without loss of generality, assume that $N = PQ$, where P and Q are integers. Define $\mathbf{V} = \frac{1}{\sqrt{Q}} \mathbf{I}_P \otimes \mathbf{1}_Q$, where $\mathbf{I}_P$ is a $P \times P$ identity matrix, $\mathbf{1}_Q$ is a $Q \times 1$ all-ones vector, and $\otimes$ denotes the Kronecker product. Denote $\mathbf{v}_p$ by the p-th column of $\mathbf{V}$, where it is easy to show that $\mathbf{v}_p^H \mathbf{v}_l = 0$ for $p \neq l$, and $\mathbf{v}_p^H \mathbf{v}_p = 1$. $\boldsymbol{\theta}_k$ is selected based on the following criterion:

$$\max_{\mathbf{v}_p} \ \frac{|\mathbf{v}_p^H \mathbf{D}_{k'} \mathbf{h}_k|^2 \alpha_2^2}{|\mathbf{v}_p^H \mathbf{D}_{k'} \mathbf{h}_k|^2 \alpha_1^2 + \sum_{i=1, i\neq k}^{K} |\mathbf{v}_p^H \mathbf{D}_{k'} \mathbf{h}_i|^2 + \frac{1}{\rho}}. \tag{6}$$

As shown in the remainder of the letter, the use of on-off control not only yields better performance than the DFT-based

[1] We note that it is possible to extend the proposed IRS-NOMA transmission scheme to the case with multiple cell-edge users served simultaneously on a single beam, which is beyond the scope of this letter.

[2] These assumptions facilitate the illustration for the benefit of IRS-NOMA. How to extend the obtained analytical results to more general scenarios is an important topic for future research but beyond the scope of this letter.

design, but also ensures that insightful analytical results can be developed. We note that both $\mathbf{D}_{k'}$ and $\mathbf{h}_k$ are complex Gaussian distributed, and the product of $\mathbf{D}_{k'}$ and $\mathbf{h}_k$ makes the analysis of the outage performance quite challenging. Therefore, in the remainder of this letter, we will focus on two special cases, one with $K = 1$ and an arbitrary choice of Q, and the other with $K \geq 2$ and $Q = 1$, where some insightful analytical results can be obtained.

Lemma 1: For the single user case ($K = 1$), the use of IRS-NOMA with on-off control can achieve the following outage probability at user k':

$$\mathrm{P}_{k'} = \frac{\xi^{\frac{N}{2}}}{(\Gamma(Q))^P}\left(\xi^{-\frac{Q}{2}}\Gamma(Q) - 2K_Q\left(2\xi^{\frac{1}{2}}\right)\right)^P, \quad (7)$$

if $\alpha_2^2 - \alpha_1^2\epsilon_{k'} > 0$, otherwise $\mathrm{P}_{k'} = 1$, where $\xi = \frac{Q\epsilon_{k'}}{\rho(\alpha_2^2 - \alpha_1^2\epsilon_{k'})}$, $\epsilon_{k'} = 2^{R_{k'}} - 1$, $R_{k'}$ denotes the target rate of user k', $K_n(\cdot)$ denotes the modified Bessel function of the second kind, and $\Gamma(\cdot)$ denotes the gamma function. At high SNR, the outage probability can be approximated as follows:

$$\mathrm{P}_{k'} \approx \begin{cases} \xi^N\left(-\ln(\xi)\right)^N, & Q = 1 \\ \frac{\xi^P}{(Q-1)^P}, & Q \geq 2 \end{cases}, \quad (8)$$

for $\alpha_2^2 - \alpha_1^2\epsilon_{k'} > 0$.

Proof: For the case $K = 1$ and $\boldsymbol{\theta}_k = \mathbf{v}_p$, $\mathrm{SINR}_{k'}$ in (2) can be simplified as follows:

$$\mathrm{SINR}_{k',p} = \frac{|\mathbf{v}_p^H\mathbf{D}_{k'}\mathbf{h}_k|^2\alpha_2^2}{|\mathbf{v}_p^H\mathbf{D}_{k'}\mathbf{h}_k|^2\alpha_1^2 + \frac{1}{\rho}}. \quad (9)$$

Because of the structure of $\mathbf{v}_p$, $\sqrt{Q}\mathbf{v}_p^H\mathbf{D}_{k'}\mathbf{h}_k$ is simply an inner product of two $Q \times 1$ complex Gaussian vectors. By first treating $\sqrt{Q}\mathbf{v}_p^H\mathbf{D}_{k'}\mathbf{h}_k$ as a complex Gaussian random variable with zero mean and variance $|\mathbf{h}_k|^2$ and using the fact that $|\mathbf{h}_k|^2$ is gamma distributed, the probability density function (pdf) of $\sqrt{Q}\mathbf{v}_p^H\mathbf{D}_{k'}\mathbf{h}_k$ can be obtained as follows:

$$f_{Q|\mathbf{v}_p^H\mathbf{D}_{k'}\mathbf{h}_k|^2}(x) = \frac{2x^{\frac{Q-1}{2}}}{\Gamma(Q)}K_{Q-1}(2\sqrt{x}), \quad (10)$$

where the details for the derivation can be found in [16], [17].

Therefore, for the case $\boldsymbol{\theta}_k = \mathbf{v}_p$, the outage probability can be expressed as follows:

$$\mathrm{P}_{k',p} = \mathrm{P}\left(\log(1 + \mathrm{SINR}_{k',p}) < R_{k'}\right). \quad (11)$$

By applying the pdf shown in (10), the outage probability can be expressed as follows:

$$\begin{aligned}\mathrm{P}_{k',p} &= \int_0^{\xi} f_{Q|\mathbf{v}_p^H\mathbf{D}_{k'}\mathbf{h}_k|^2}(x)dx \\ &= \frac{2}{\Gamma(Q)}\int_0^{\xi} x^{\frac{Q-1}{2}}K_{Q-1}(2\sqrt{x})dx \\ &= \frac{1}{\Gamma(Q)}\xi^{\frac{Q+1}{2}}\left(\xi^{-\frac{Q+1}{2}}\Gamma(Q) - 2\xi^{-\frac{1}{2}}K_Q\left(2\xi^{\frac{1}{2}}\right)\right), \end{aligned} \quad (12)$$

where the last step follows from Eq. (6.561.8) in [18].

For IRS-NOMA with on-off control, $\mathbf{V} = \frac{1}{\sqrt{Q}}\mathbf{I}_P \otimes \mathbf{1}_Q$, and hence one can easily verify that $\mathbf{v}_p^H\mathbf{D}_{k'}\mathbf{h}_k$ and $\mathbf{v}_l^H\mathbf{D}_{k'}\mathbf{h}_k$ are independent and identically distributed (i.i.d.) for $p \neq l$. Therefore, the use of the selection criterion in (6) ensures that the outage probability at user k' can be expressed as follows:

$$\mathrm{P}_{k'} = \frac{\xi^{\frac{P(Q+1)}{2}}}{(\Gamma(Q))^P}\left(\xi^{-\frac{Q+1}{2}}\Gamma(Q) - 2\xi^{-\frac{1}{2}}K_Q\left(2\xi^{\frac{1}{2}}\right)\right)^P. \quad (13)$$

With some algebraic manipulations, (7) in the lemma can be obtained.

In order to find the high SNR approximation for (7), we first note that at high SNR, $\rho \to \infty$, which means that $\xi \to 0$. Recall that $K_n(z)$ can be approximated as follows: [18]

$$K_n(z) \approx \frac{1}{2}\left(\frac{(n-1)!}{\left(\frac{z}{2}\right)^n} - \frac{(n-2)!}{\left(\frac{z}{2}\right)^{n-2}}\right), \quad (14)$$

for $n \geq 2$ and $z \to 0$. Therefore, the outage probability $\mathrm{P}_{k'}$ can be approximated as follows:

$$\begin{aligned}\mathrm{P}_{k'} &= \frac{1}{(\Gamma(Q))^P}\left(\Gamma(Q) - 2\xi^{\frac{Q}{2}}K_Q\left(2\xi^{\frac{1}{2}}\right)\right)^P \\ &\approx \frac{1}{(\Gamma(Q))^P}\left(\Gamma(Q) - \xi^{\frac{Q}{2}}\left(\frac{(Q-1)!}{\xi^{\frac{Q}{2}}} - \frac{(Q-2)!}{\xi^{\frac{Q-2}{2}}}\right)\right)^P \\ &\approx \frac{\xi^P}{(Q-1)^P}, \end{aligned} \quad (15)$$

for the cases with $Q \geq 2$.

For the case with $Q = 1$, unlike (14), a different approximation for the Bessel function will be used as shown in the following:

$$K_1(z) \approx \frac{1}{2}\frac{1}{\left(\frac{z}{2}\right)} + \left(\frac{z}{2}\right)\ln\left(\frac{z}{2}\right), \quad (16)$$

for $z \to 0$. Therefore, for the case with $Q = 1$, the outage probability can be approximated as follows:

$$\begin{aligned}\mathrm{P}_{k'} &= \left(1 - 2\xi^{\frac{1}{2}}K_1\left(2\xi^{\frac{1}{2}}\right)\right)^N \\ &\approx \left(1 - \xi^{\frac{1}{2}}\left(\frac{1}{\xi^{\frac{1}{2}}} + \xi^{\frac{1}{2}}\ln(\xi)\right)\right)^P \approx \xi^N\left(-\ln(\xi)\right)^N. \end{aligned} \quad (17)$$

By combining (15) with (17), (8) in the lemma can be obtained, and the proof for the lemma is complete. □

Remark 1: The diversity gain for the case with $Q = 1$ can be found as follows:

$$\begin{aligned}-\lim_{\rho\to\infty}\frac{\log \mathrm{P}_{k'}}{\log\rho} &= \lim_{\xi\to 0}\frac{\log[\xi^N\left(-\ln(\xi)\right)^N]}{\log\xi} \\ &= N + N\lim_{\xi\to 0}\frac{\log[(-\ln\xi)]}{\log\xi} = N, \end{aligned} \quad (18)$$

where the last step follows by applying L'Hospital's rule. It is straightforward to show that the diversity gain for the case $Q \geq 2$ is P. Therefore, the choice of $Q = 1$ is diversity optimal to IRS-NOMA with on-off control.

Remark 2: Lemma 1 is only applicable to IRS-NOMA with on-off control. The analytical results for IRS-NOMA with DFT are difficult to obtain, mainly because of the correlation between $|\mathbf{v}_p^H\mathbf{D}_{k'}\mathbf{h}_k|^2$ and $|\mathbf{v}_i^H\mathbf{D}_{k'}\mathbf{h}_k|^2$, for $i \neq p$. We note that simulation results indicate that this correlation is very weak, which results in an observation that the diversity order achieved by the DFT case is similar to that of the scheme with on-off control.

Note that for the multi-user case ($K \geq 2$), the outage probability achieved by IRS-NOMA with on-off control is difficult to analyze, due to the correlation between $|\mathbf{v}_p^H \mathbf{D}_{k'} \mathbf{h}_k|^2$ and $|\mathbf{v}_p^H \mathbf{D}_{k'} \mathbf{h}_i|^2$, for $i \neq p$. Consistent to the single-user case, simulation results show that $Q = 1$ is also optimal in the high SNR regime for the multi-user case. Therefore, the choice of $Q = 1$ is focused in the following, where a closed-form expression for the outage probability can be obtained, as shown in the the following lemma.

Lemma 2: For the multi-user case $K \geq 2$, the outage probability achieved by IRS-NOMA with on-off control ($Q = 1$) is given by

$$\mathrm{P}_{k'} = \left(1 - \frac{2\sqrt{\frac{\epsilon_{k'}}{\rho\tau}} K_1\left(2\sqrt{\frac{\epsilon_{k'}}{\rho\tau}}\right)}{\left(1 + \frac{\epsilon_{k'}}{\tau}\right)^{K-1}}\right)^N, \tag{19}$$

where $\tau = \alpha_2^2 - \epsilon_{k'}\alpha_1^2$. At high SNR, the outage probability can be approximated as follows:

$$\mathrm{P}_{k'} \approx \left(1 - \frac{1}{\left(1 + \frac{\epsilon_{k'}}{\tau}\right)^{K-1}}\right)^N. \tag{20}$$

Proof: For the special case $Q = 1$, $\mathbf{v}_p$ is an $N \times 1$ vector with all of its elements being zero except its p-th element being one. Therefore, $\mathbf{v}_p^H \mathbf{D}_{k'} \mathbf{h}_k$ becomes $d_{k',p} h_{k,p}$, where $d_{k',p}$ is the p-th element on the diagonal of $\mathbf{D}_{k'}$ and $h_{k,p}$ is the p-th element of $\mathbf{h}_k$. Therefore, with $\boldsymbol{\theta}_k = \mathbf{v}_p$, $\mathrm{SINR}_{k'}$ can be simplified as follows:

$$\mathrm{SINR}_{k',p} = \frac{|d_{k',p}|^2 |h_{k,p}|^2 \alpha_2^2}{|d_{k',p}|^2 |h_{k,p}k|^2 \alpha_1^2 + |d_{k',p}|^2 \sum_{i=1, i\neq k}^{K} |h_{i,p}|^2 + \frac{1}{\rho}}.$$

We first note that $\mathbf{h}_k = \mathbf{G}_k \mathbf{w}_k$ is still a complex Gaussian vector, since $\mathbf{w}_k$ is normalized. We further note that $\mathbf{h}_k$ and $\mathbf{h}_i$, $k \neq i$, are independent since $\mathbf{w}_k$ and $\mathbf{w}_i$ are assumed to be orthonormal vectors. Therefore, $h_{k,p}$ and $h_{i,p}$ are independent and complex Gaussian distributed. Hence, the SINR can be further simplified as follows:

$$\mathrm{SINR}_{1',p} = \frac{xy\alpha_2^2}{xy\alpha_1^2 + xz + \frac{1}{\rho}}, \tag{21}$$

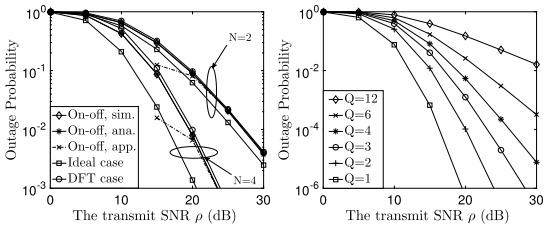
where $x = |d_{k',p}|^2$ and $y = |h_{k,p}|^2$ are two independent and exponentially distributed random variables, and $z = \sum_{i=1, i\neq k}^{K} |h_{i,p}|^2$ is gamma distributed. Therefore, the outage probability can be expressed as follows:

$$\begin{aligned}\mathrm{P}_{k',p} &= \int_0^\infty \int_0^\infty \left(1 - e^{-\frac{\epsilon_{k'} xz + \frac{\epsilon_{k'}}{\rho}}{x(\alpha_2^2 - \epsilon_{k'}\alpha_1^2)}}\right) \frac{e^{-x} dx}{(K-2)!} z^{K-2} e^{-z} dz \\ &= 1 - \frac{1}{(K-2)!} 2\sqrt{\frac{\epsilon_{k'}}{\rho\tau}} K_1\left(2\sqrt{\frac{\epsilon_{k'}}{\rho\tau}}\right) \\ &\quad \times \int_0^\infty e^{-\frac{\epsilon_{k'}}{\tau} z} z^{K-2} e^{-z} dz,\end{aligned}$$

where the last step follows from Eq. (3.324.1) in [18].

By applying Eq. (3.381.4) in [18], the outage probability when $\boldsymbol{\theta}_k = \mathbf{v}_p$ can be obtained as follows:

$$\mathrm{P}_{k',p} = 1 - 2\sqrt{\frac{\epsilon_{k'}}{\rho\tau}} K_1\left(2\sqrt{\frac{\epsilon_{k'}}{\rho\tau}}\right) \frac{1}{\left(1 + \frac{\epsilon_{k'}}{\tau}\right)^{K-1}}. \tag{22}$$



(a) Comparison between different IRS-NOMA ($Q = 1$) (b) Impact of the choice of Q on IRS-NOMA ($N = 12$)

Fig. 2. Impact of IRS-NOMA on the downlink outage probability for the single-user case ($K = 1$). $M = 4$. $R_{k'} = 2$ bits per channel use (BPCU).

Because of the structure of $\mathbf{V}$, the SINRs for different $\mathbf{v}_p$ are i.i.d., and therefore, the outage probability $\mathrm{P}_{k'}$ can be obtained as shown in (19) in the lemma.

At high SNR, by using the fact that $K_1(x) \approx \frac{1}{x}$ for $x \to 0$, the outage probability can be approximated as follows:

$$\mathrm{P}_{k'} \approx \left(1 - \frac{2\sqrt{\frac{\epsilon_{k'}}{\rho\tau}} \left(2\sqrt{\frac{\epsilon_{k'}}{\rho\tau}}\right)^{-1}}{\left(1 + \frac{\epsilon_{k'}}{\tau}\right)^{K-1}}\right)^N.$$

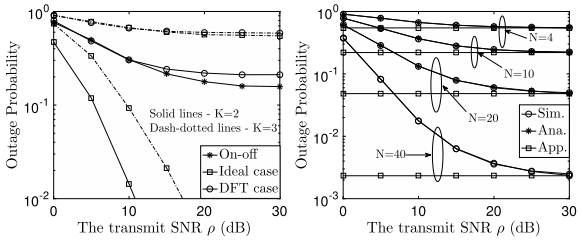
With some algebraic manipulations, the approximation shown in (20) can be obtained, and the lemma is proved. □

Remark 3: The high SNR approximation shown in (20) indicates the existence of an error floor for the outage probability, i.e., the outage probability does not go to zero by simply increasing the transmission power. However, the outage probability can be reduced by increasing N, as shown in (20).

IV. Numerical Results

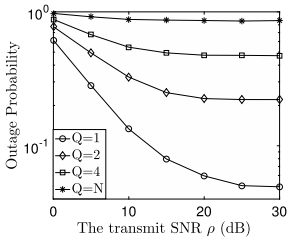
In this section, computer simulation results are presented to demonstrate the performance of IRS-NOMA, where we use $\alpha_1^2 = \frac{1}{5}$ and $\alpha_2^2 = \frac{4}{5}$.Since it is assumed that there is no direct link between the IRS's and the near users, the near users' performance is exactly the same as those in conventional NOMA systems, and hence we focus on the cell-edge users' performance. In Fig. 2, the performance of IRS-NOMA is studied by focusing on the single-user case ($K = 1$). Fig. 2(a) shows that the slop of the outage probability curves for the three schemes is the same, which indicates that they achieve the same diversity order. Among the three schemes, the one with ideal beamforming yields the best performance, but it might not be supported by a practical antenna array. Among the two practical IRS-NOMA schemes, the one with on-off control yields better performance. Fig. 2(a) also confirms the accuracy of the developed analytical results shown in Lemma 1. Remark 1 indicates that increasing Q decreases the achieved diversity gain, which is confirmed by Fig. 2(b).

Fig. 3 shows the performance of the IRS-NOMA schemes when there are multiple users ($K \geq 2$). Fig. 3(a) shows that for IRS-NOMA with ideal beamforming, the outage probability can be reduced to zero by increasing the transmission power. However, there are error floors for the two practical IRS-NOMA schemes. The reason for these error floors is due to the fact that the use of finite-resolution beamforming cannot eliminate inter-pair interference completely. Fig. 3(a) also shows that the on-off scheme outperforms the DFT one, which is consistent to Fig. 2(a). The accuracy of the analytical results shown in Lemma 2 is confirmed by Fig. 3(b). In addition,



(a) Comparison between different IRS-NOMA ($N = 4$ and $Q = 1$)

(b) Effects of N ($Q = 1$)



(c) Effects of Q ($N = 20$)

Fig. 3. Impact of IRS-NOMA on the downlink outage probability for the multi-user case ($K \geq 2$). $M = 4$. $R_{k'} = 1$ BPCU.

Fig. 3(b) also shows that increasing N can effectively reduce the outage probability, as indicated in (20) in Lemma 2. Fig. 3(c) confirms the optimality of the choice of $Q = 1$ in the multi-user scenario, which is consistent to Fig. 2(b).

Finally, the impact of path loss on the performance of IRS transmission is studied in Fig. 4, where conventional relaying is used as a benchmarking scheme. The distance between the base station and the IRS is 20 m, and the distance between the IRS and user k' is 10 m. For conventional relaying, the location of the relay is assumed to be the same as that of the IRS. The path loss exponent is 3.5 and the noise power is -70 dBm. It is assumed that the base station and the conventional relay use the same transmission power. An important observation from the figure is that conventional relaying can outperform IRS transmission when the transmission power is small. This performance loss is due to the severe path loss suffered by IRS transmission. However, by increasing the transmission power or the number of elements on the IRS, IRS transmission can realize a significant performance gain over conventional relaying, as can be observed from the figure.

V. Conclusion

In this letter, IRS-NOMA transmission has been proposed to ensure that more users can be served on each orthogonal spatial direction, compared to SDMA. In addition, the impact of hardware impairments on the design of IRS-NOMA has been investigated and the performance of practical IRS-NOMA transmission has also been characterized. In this letter, it was assumed that there is no direct link between the base station and a cell-edge user, and we note that the proposed IRS-NOMA scheme can be extended to the case in which there is a direct link between a cell-edge user and the base station. With these direct links, there are two possible beamforming designs. One is to simultaneously serve multiple near and far users on each beam, where the beams can be orthogonal to each other if the number of beams is less than the number of antennas at the base station. The other is to serve a single user on each beam, which means that $2K$ beams are needed for serving $2K$ users and these beams have to be non-orthogonal if $2K > M$ [19].

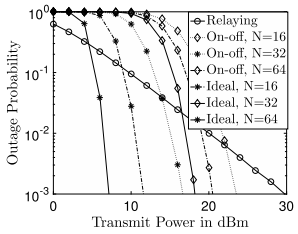


Fig. 4. Impact of path loss on the performance of IRS transmission. $K = 1$, $Q = 1$, $M = 4$, and $R_{k'} = 2$ BPCU.

References

[1] Z. Ding *et al.*, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[2] *Study on Downlink Multiuser Superposition Transmission for LTE*, 3rd Generation Partnership Project (3GPP), Mar. 2015.

[3] *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 15)*, 3rd Generation Partnership Project (3GPP), Jan. 2019.

[4] *Study on Non-Orthogonal Multiple Access (NOMA) for NR (Release 16)*, 3rd Generation Partnership Project (3GPP), Dec. 2018.

[5] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6174–6189, Dec. 2016.

[6] M. D. Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *J. Wireless Com Netw.*, vol. 2019, no. 1, pp. 1–20, Dec. 2019.

[7] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[8] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[9] C. Huang, G. C. Alexandropoulos, A. Zappone, M. Debbah, and C. Yuen, "Energy efficient multi-user MISO communication using low resolution large intelligent surfaces," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, UAE, Dec. 2018, pp. 1–6.

[10] Q. Wu and R. Zhang, "Beamforming optimization for intelligent reflecting surface with discrete phase shifts," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 7830–7833.

[11] C. You, Y. Zeng, R. Zhang, and K. Huang, "Asynchronous mobile-edge computation offloading: Energy-efficient resource management," 2018, *arXiv:1801.03668*. [Online]. Available: http://arxiv.org/abs/1801.03668

[12] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications," 2019, *arXiv:1911.03084*. [Online]. Available: http://arxiv.org/abs/1911.03084

[13] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," 2019, *arXiv:1909.03272*. [Online]. Available: http://arxiv.org/abs/1909.03272

[14] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[15] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557–1571, Apr. 2016.

[16] N. C. Sagias, "On the ASEP of decode-and-forward dual-hop networks with pilot-symbol assisted M-PSK," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 510–521, Feb. 2014.

[17] H. Liu, H. Ding, L. Xiang, J. Yuan, and L. Zheng, "Outage and BER performance analysis of cascade channel in relay networks," *Procedia Comput. Sci.*, vol. 34, pp. 23–30, 2014.

[18] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. New York, NY, USA: Academic, 2000.

[19] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-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. 2016.