NOMA Meets Finite Resolution Analog Beamforming in Massive MIMO and Millimeter-Wave Networks

Zhiguo Ding, Senior Member, IEEE, Linglong Dai, Senior Member, IEEE, Robert Schober, Fellow, IEEE, and H. Vincent Poor, Fellow, IEEE

Abstract—Finite resolution analog beamforming (FRAB) has been recognized as an effective approach to reduce hardware costs in massive multiple-input multiple-output (MIMO) and millimeter-wave networks. However, the use of FRAB means that the beamformers are not perfectly aligned with the users’ channels and multiple users may be assigned similar or even identical beamformers. This letter shows how non-orthogonal multiple access (NOMA) can be used to exploit this feature of FRAB, in which a single FRAB-based beamformer is shared by multiple users. Both analytical and simulation results are provided to demonstrate the excellent performance achieved by this new NOMA transmission scheme.

Index Terms—Non-orthogonal multiple access (NOMA), massive MIMO, millimeter-wave (mmWave) networks, finite resolution analog beamforming (FRAB).

I. INTRODUCTION

NON-ORTHOGONAL multiple access (NOMA) is a promising multiple access technique for next generation wireless networks [1]–[3], and has been shown to be compatible with many important 5G technologies, including massive multiple-input multiple-output (MIMO) and millimeter wave (mmWave) transmission [4]–[6].

A recent development in massive MIMO and mmWave networks is the use of finite resolution analog beamforming (FRAB), which reduces hardware costs significantly compared to purely digital precoding [7], [8]. Analog beamforming does not alter the amplitude of a signal, but modifies its phase only, which is different from digital beamforming. Analog beamforming can be either used alone by a transmitter or combined with conventional digital beamforming, which results in so-called hybrid precoding. The finite resolution constraint on analog beamforming is due to the fact that the number of phase shifts supported by a practical circuit is finite [7], [9]. An example for one-bit resolution analog beamforming is provided in Table 1, where $S_1$ and $S_2$ denote two distinct subsets of users to be described in more detail below. Depending on the values of a user’s complex-valued channel coefficients, 1 or −1 will be chosen as the beamformer coefficients, as shown in Table I. The reduced hardware costs of FRAB are at the expense of performance losses since the obtained beamformers are not perfectly aligned with the target users’ channels.

The purpose of this letter is to demonstrate that the characteristics of FRAB favor the use of NOMA. Consider again the example shown in Table 1. To clearly illustrate the benefits of the combination of FRAB and NOMA, the users’ channel vectors are chosen to be orthogonal to each other. Assume that the users in $S_1$ need to be served first because of their high priority traffic. If digital beamforming with perfect resolution were used, the beamforming vector for user 1 in $S_1$ would be simply this user’s channel vector, and therefore this beamformer could not be used by the two users in $S_2$, since this beamformer would be orthogonal to these two users’ channel vectors. On the other hand, if FRAB is used, the beamformer preferred by user $i$ in $S_1$ is exactly the same as that of user $i$ in $S_2$, even though the two users have orthogonal channel vectors. As a result, the use of NOMA ensures that all four users can communicate concurrently.

In this letter, a new NOMA transmission scheme that exploits the features of FRAB is proposed, and analytical results for the corresponding outage probabilities and diversity gains of the users are presented. The provided simulation results demonstrate not only the excellent performance of the proposed NOMA scheme, but also the accuracy of the developed analytical results. We note that the developed analytical results concerning the diversity gains are also applicable to conventional MIMO scenarios without NOMA, and hence shed light on the performance loss caused by FRAB in a general MIMO network.

II. SYSTEM MODEL

Consider a NOMA downlink scenario, in which the base station is equipped with $M$ antennas. Assume that there are two groups of single-antenna users in the network. Denote by $S_1$ a set of users with strict quality of service (QoS)}
requirements, whose distances to the base station are denoted by $d_{yk}$ and are assumed to be fixed. Denote by $S_2$ a set of users to be served opportunistically, and these users are uniformly distributed in a disk-shaped area with radius $r_1$, where the base station is at the center. Denote the distances of the users in $S_2$ to the base station by $d_{yi}$. The $M \times 1$ channel vector of a user in $S_1$ ($S_2$) is denoted by $\mathbf{h}_i$ ($\mathbf{g}_i$). Two types of channel models are considered in this letter, namely Rayleigh fading which is commonly used to model massive MIMO channels, and the mmWave model [8]. The mmWave channel vector is modeled as follows:

$$\mathbf{h}_i = \frac{d_k}{1 + d_{yk}^2} \left[ 1 + e^{-j \pi \theta_k} \ldots e^{-j \pi (M-1) \theta_k} \right]^T. \tag{1}$$

Here, $\alpha$ denotes the path loss exponent, $\theta_k$ is the normalized direction, and $a_k$ denotes the fading attenuation coefficient. Note that for the purpose of illustration, only the line-of-sight path is considered for the mmWave model.

### A. Implementation of Finite Resolution Analog Beamforming

Suppose that the users in $S_1$ are served via FRAB. Denote by $\mathbf{f}_i$ the $M \times 1$ beamforming vector for user $k$, where each element of $\mathbf{f}_i$ is drawn from the following vector:

$$\mathbf{f} = \left[ 1 e^{j \frac{2 \pi}{N_q}} \ldots e^{j \frac{(N_q-1) \pi}{N_q}} \right], \tag{2}$$

where $N_q$ denotes the number of supported phase shifts.

The $i$-th element of $\mathbf{f}$ is chosen as the $m$-th element of $\mathbf{f}_i$ based on the following criterion:

$$i_{k,m}^* = \arg \min_{i \in \{1, \ldots, N_q\}} \left| \frac{f_i - h_{km}}{h_{km}} \right|^2, \tag{3}$$

where $f_i$ denotes the $i$-th element of $\mathbf{f}$, and $h_{km}$ denotes the $m$-th element of user $k$’s channel vector.

### B. Implementation of NOMA

To reduce the system complexity, suppose that only one user from $S_2$ will be paired with user $k$ from $S_1$ and denote this user by user $i_k^*$. The base station sends a superposition of the messages of the two users on each beam. User $k$ in $S_1$ treats its partner’s message as noise and decodes its own message with the following signal-to-interference-plus-noise ratio (SINR):

$$\text{SINR}_k = \frac{|\mathbf{h}_k^H \mathbf{f}_k|^2 a_{1,k}^2}{|\mathbf{h}_k^H \mathbf{f}_k|^2 a_{1,k}^2 + \sum_{l \in S_1 \backslash k} |\mathbf{h}_k^H \mathbf{f}_l|^2 + \frac{M}{\rho}}. \tag{4}$$

where the factor $M$ in the denominator is due to the transmit power normalization, and the power allocation coefficients are denoted by $a_{n,k}$. Note that $\sum_{n=0}^{i} a_{n,k}^2 = 1$ and $a_{0,k} \geq a_{1,k}$.

By applying successive interference cancellation (SIC), user $i_k^*$ can decode its partner’s message with the following SINR:

$$\text{SINR}_{k_{i_k^*}} = \frac{|\mathbf{g}_{i_k^*}^H \mathbf{f}_{i_k^*}|^2 a_{1,k}^2}{|\mathbf{g}_{i_k^*}^H \mathbf{f}_{i_k^*}|^2 a_{1,k}^2 + \sum_{l \in S_1 \backslash k} |\mathbf{g}_{i_k^*}^H \mathbf{f}_l|^2 + \frac{M}{\rho}}. \tag{5}$$

We use the following user selection criterion:

$$i_{k}^* = \arg \max \{ \text{SINR}_{k_{i_k^*}} \ldots \text{SINR}_{k_{i_k^*}} \}. \tag{6}$$

Note that this criterion selects that user which maximizes the probability of successful intra-NOMA interference cancellation, a key stage for SIC. Since the users in $S_2$ are served opportunistically, we allow one user from $S_2$ to be included in more than one pair. More sophisticated user scheduling algorithms can be designed to realize fairness for the users in $S_2$; these are not presented here due to space limitations.

### III. Performance Analysis

To the best of the authors’ knowledge, the impact of FRAB on the diversity gain has not been analyzed yet, not even for scenarios without NOMA. In order to obtain insight into the performance of the proposed NOMA scheme, in this section, we focus on the special case of $N_q = 2$, $|S_1| = 1$, and Rayleigh fading channels. Note that $N_q = 2$ represents the case of one-bit resolution analog beamforming [9].

#### A. Performance of the User in $S_1$

When there is a single beam, i.e., $|S_1| = 1$, the outage probability achieved by the user in $S_1$ is given by

$$P_k^o = P \left( |\mathbf{h}_k^H \mathbf{f}_k|^2 < \phi_1 \right),$$

where $\phi_1 = \frac{\alpha_1^2}{\alpha_1^2 - \epsilon_i \sum_{n=1}^{M} a_{n,k}^2}$, $i \in \{0, 1\}$. Note that $\alpha_1^2 > \epsilon_i \sum_{n=1}^{M} a_{n,k}^2$ is assumed in this letter, since otherwise the outage probability is always one. In order to find the cumulative distribution function (CDF) of $|\mathbf{h}_k^H \mathbf{f}_k|^2$, the following proposition is provided first.

**Proposition 1:** Consider $M$ independent and identically distributed (i.i.d.) random variables, denoted by $z_m$, each of which follows the folded normal distribution, i.e., $z_m$ is the absolute value of a Gaussian variable with mean 0 and variance $\frac{1}{2}$. The CDF of $z_m$ can be approximated as follows:

$$F_{z_m}(z) \approx \frac{2^M z^{\frac{M}{2}}}{\pi^{\frac{M}{2}} M!}, \tag{7}$$

when $z \rightarrow 0$.

The following lemma provides an asymptotic approximation for the CDF of the effective channel gains of the user in $S_1$.

**Lemma 1:** For user $k$ in $S_1$, the CDF of its effective channel gain on beam $\mathbf{f}_k$ can be approximated as follows:

$$F_{|\mathbf{h}_k^H \mathbf{f}_k|^2}(y) \approx \frac{2^M \left[ y(1 + d_{yk}^a) \right]^{\frac{M+1}{2}} B \left( \frac{3}{2}, \frac{M}{2} \right)}{\pi^{\frac{M}{2}} (M - 1)! M^2 \Gamma \left( \frac{3}{2} \right)}, \tag{8}$$

for $y \rightarrow 0$, where $B(\cdot)$ denotes the beta function and $\Gamma(\cdot)$ denotes the gamma function.

By using Lemma 1 and with some algebraic manipulations, the following corollary can be obtained.

**Corollary 1:** For the proposed NOMA system with Rayleigh fading, user $k$ from $S_1$ achieves a diversity gain of $\frac{M+1}{2}$.

---

1 Since $|S_1| = 1$, there is only one user in $S_1$. To keep the notation consistent with that in the previous section, this user is denoted by user $k$ in the remainder of this section.
Remark 1: With \( M \) antennas at the base station, the full diversity gain for the considered scenario is \( M \), but only a diversity gain of \( \frac{M+1}{2} \) is achieved by the proposed scheme. This performance loss is mainly due to the use of FRAB.

Remark 2: It is important to point out that Corollary 1 is general and applicable to conventional MIMO networks without NOMA. In Fig. 1, two types of channel correlations. The NOMA power coefficients are set as \( N_0 = 2, r_1 = 40m, r_2 = r_1, \alpha = 3 \), \( R_0 = 1 \) bit per channel use (BPCU), and \( R_1 = 1.5 \) BPCU.

B. Performance of the Paired User in \( S_2 \)

When there is a single beam, the outage probability achieved by the user in \( S_2 \) paired with user \( k \) is \( \mathcal{P}_k \) becomes \( \pi \mathcal{P}_k \) = 
\[ 1 - P \left( |g_{ik}|^2 > \max(\phi_0, \phi_1) \right) \] . For the case of \( |S_1| = 1 \), the proposed user selection criterion shown in (6) simplifies to
\[ i_k^* = \arg\max \{ |g_{i_k}^H f_k|^2, \ldots, |g_{S_2}^H f_k|^2 \}, \]
which is a monotonically increasing function of \( x \). Therefore, the CDF of \( |g_{i_k}^H f_k|^2 \) can be obtained as
\[ F_\pi(x) = \sum_{i=1}^{N} w_i (1 - e^{-c_i x}), \]
where \( N \) is a parameter for the Gauss-Chebyshev approximation, \( w_i = \frac{x_i}{N} \), and \( c_i = 1 + (\frac{4}{x_i})^2 \).

After applying the simplified criterion in (9) and assuming that the users’ channels are i.i.d., the outage probability of user \( i_k^* \) for decoding its message delivered on beam \( k \) is
\[ \left( F_\pi(x, \phi_0, \phi_1) \right)^{|S_1|}. \]

Corollary 2: For the proposed NOMA system with Rayleigh fading, the full diversity gain of \( |S_2| \) is achievable by user \( i_k^* \) in \( S_2 \).

IV. NUMERICAL RESULTS

In this section, the performance of the proposed NOMA-MIMO scheme is evaluated by using computer simulations. The NOMA power coefficients are set as \( \alpha_{0,k} = 3/4 \) and \( \alpha_{1,k} = 1/4 \). The noise power is \(-30dBm\). Without loss of generality, assume that the users in \( S_1 \) lie on a circle with radius \( r_1 \), and the base station is located at its center.

Fig. 1. Outage rates achieved by the considered schemes for different channel models. \( M = 30, |S_1| = 3, |S_2| = 500, N_0 = 2, r_1 = 40m, r_2 = r_1, \alpha = 3 \), \( R_0 = 1 \) bit per channel use (BPCU), and \( R_1 = 1.5 \) BPCU.

are observed from the figure, the use of NOMA can result in a significant performance gain compared to the scheme without NOMA. For example, when the transmission power of the base station is 15dBm, the use of NOMA can offer a rate improvement of 3 bits per channel use (BPCU) over the conventional MIMO scheme, for both considered channel models. In Fig. 2, we focus on the special case \( |S_1| = 1 \), where the diversity gain achieved by NOMA is studied. To facilitate this diversity analysis, we set \( |S_2| = M \), which means that the diversity gains for the users in \( S_1 \) and \( S_2 \) are \( \frac{M+1}{2} \) and \( M \), respectively. From the figure, one can clearly observe the loss of diversity gain for the user from \( S_1 \) due to FRAB. Note that the curves for the analytical results match those of the simulations, which verifies the accuracy of our analysis.

V. CONCLUSIONS

In this letter, NOMA has been proposed as a means of mitigating the reduced degrees of freedom induced by FRAB in massive MIMO and mmWave networks. The developed analytical and simulation results have demonstrated the superior performance of the proposed NOMA scheme.

APPENDIX A

PROOF OF PROPOSITION 1

Recall that the probability density function (pdf) of a random variable generated by folding a normally distributed variable with zero mean and variance one is \( f_{z_m}(x) = \frac{2}{\sqrt{\pi}} e^{-x^2} \). Therefore, the CDF of the square of the sum of \( z_m \) is given by
\[ F_{z_m} = \frac{2^M}{\pi^M} \prod_{m=1}^{M} \int \prod_{m=1}^{M} e^{-x^2} dx_m, \]
(11)
where the subscript \( M \) is added to show that the CDF is a function of \( M \), and to facilitate the following analysis.

The proposition can be proved by using the inductive method. For the case \( M = 1 \), the approximate expression in the proposition can be simplified as follows:

\[
F_{\tilde{Z}_1}(z) \approx \frac{2}{\pi} \left( \frac{z}{1 - z} \right). \tag{12}
\]

By calculating the integral of the pdf of \( z_m \) and applying \( e^{-x} \approx 1 - x \) for \( x \to 0 \), one can verify that (12) is a valid approximate expression for the CDF.

Assuming that the approximation is correct for the case of \( M = n \), \( F_{\tilde{Z}_{n+1}}(z) \) can be expressed as follows:

\[
F_{\tilde{Z}_{n+1}}(z) = \frac{2^{n-1}}{\pi} \int_0^{\sqrt{n}} \left( \sum_{m=1}^{n} e^{-\frac{z^2}{n-m+1}} \right) dz.
\]

Therefore, the approximation expression is correct for \( M = n + 1 \), and the proof is complete via induction.

**APPENDIX B**

**PROOF FOR LEMMA 1**

With FRAB, the user's effective channel gain can be expressed as follows:

\[
|h_k^H f_t|^2 = \sum_{m=1}^{M} h_{k,m}^2 e^{-j\Delta_{k,m}}. \tag{15}
\]

When \( N_q = 2 \), we separate the real and imaginary parts of the user's channel as follows:

\[
|h_k^H f_t|^2 = \sum_{m=1}^{M} (h_{k,m,real}^2 + jh_{k,m,imag}^2)^2 = \sum_{m=1}^{M} (h_{k,m,real}^2 + j\text{sign}(h_{k,m,real})h_{k,m,imag}^2)^2.
\]

\[
\text{sign}(\cdot) \text{ is the sign operation. This separation leads to the following expression:}
\]

\[
|h_k^H f_t|^2 = \sum_{m=1}^{M} \tilde{h}_m^2 + |\tilde{h}_0|^2, \tag{17}
\]

where \( \tilde{h}_0 = \sum_{m=1}^{M} \text{sign}(h_{k,m,real})h_{k,m,imag} \) and \( \tilde{h}_m = |h_{k,m,real}| \). Note that \( \text{sign}(h_{k,m,real}) \) is independent of \( h_{k,m,real} \), because the phase and the amplitude of a complex Gaussian random variable are independent. Therefore, \( \sum_{m=1}^{M} |\tilde{h}_m|^2 \) and \( |\tilde{h}_0|^2 \) are independent.

Define \( z_0 = |\tilde{h}_0|^2 \) whose pdf can be easily obtained as follows: Recall that the sum of i.i.d. Gaussian random variables is still a Gaussian variable, which means \( h_0 \) is Gaussian with zero mean and variance \( \frac{M}{M+1} \). Therefore, the CDF of \( z_0 \) is \( F_{z_0}(z) = \frac{\gamma(\frac{1}{2})}{\Gamma(\frac{1}{2})} \), and the pdf is \( f_{z_0}(z) = \frac{dF_{z_0}(z)}{dz} \), where \( \gamma(x, y) \) denotes the incomplete gamma function. Note that when \( x \to 0 \), this CDF can be approximated as follows:

\[
F_{z_0}(z) \approx \frac{2(1 + d_{y}^n)^{\frac{1}{2}} z^{\frac{1}{2}}}{M + 1}\Gamma(\frac{1}{2}). \tag{18}
\]

Therefore, the CDF of \( |h_k^H f_t|^2 \) can be expressed as follows:

\[
F_{|h_k^H f_t|^2}(y) = \int_{x+y}^{x} f_{z_0}(z) f_{z_0}(z) dz dx.
\]

According to Proposition 1, when \( z \to 0 \), the pdf of \( \tilde{Z}_M \) can be approximated as follows:

\[
f_{\tilde{Z}_M}(x) \approx \frac{2^{M-1}(1 + d_{y}^n)^{\frac{1}{2}} x^{M-2}}{\pi^{\frac{M}{2}} (M-1)!}.
\]

Therefore, when \( y \to 0 \), we have the following approximation:

\[
F_{|h_k^H f_t|^2}(y) \approx \int_{0}^{y} 2^{M-1}(1 + d_{y}^n)^{\frac{1}{2}} x^{M-2} F_{z_0}(y-x) dx,
\]

which yields the following approximation:

\[
F_{|h_k^H f_t|^2}(y) \approx \frac{2^{M-1}(1 + d_{y}^n)^{\frac{1}{2}} y^{M-2}}{\pi^{\frac{M}{2}} (M-1)!} F_{z_0}(y-x) dx.
\]

Using the definition of the beta function in the above expression, the lemma is proved.

**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] L. Song, Y. Li, Z. Ding, and H. V. Poor, “Resource management in non-orthogonal multiple access networks for 5G and beyond,” IEEE Netw., to be published.

[3] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, “Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges,” IEEE Commun. Surveys Tuts., to be published.

[4] Z. Ding and H. V. Poor, “Design of massive-MIMO-NOMA with limited feedback,” IEEE Signal Process. Lett., vol. 23, no. 5, pp. 629–633, May 2016.

[5] X. Liu and X. Wang, “Efficient antenna selection and user scheduling in 5G massive MIMO-NOMA system,” in Proc. IEEE Veh. Tech. Conf., Nanjing, China, May 2016, pp. 1–5.

[6] Z. Ding, P. Fan, and H. V. Poor, “Random beamforming in millimeter-wave NOMA networks,” IEEE Access, to be published.

[7] A. Alkhateeb, Y.-H. Nam, J. Zhang, and R. W. Heath, “Massive MIMO combining with switches,” IEEE Wireless Commun. Lett., vol. 5, no. 3, pp. 232–235, Jun. 2016.

[8] X. Gao, L. Dai, Y. Sun, S. Han, and C.-L. I., “Machine learning inspired energy-efficient hybrid precoding for mmWave massive MIMO systems,” in Proc. IEEE Int. Conf. Commun., Paris, France, May 2017.

[9] H. Yang et al., “A 1-bit 10 × 10 reconfigurable reflectarray antenna: Design, optimization, and experiment,” IEEE Trans. Antennas Propag., vol. 64, no. 6, pp. 2246–2254, Jun. 2016.

[10] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, “On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users,” IEEE Signal Process. Lett., vol. 21, no. 12, pp. 1501–1505, Dec. 2014.