

# Two-Stage Analog Combining in Hybrid Beamforming Systems with Low-Resolution ADCs

Jinseok Choi, *Student Member, IEEE*, Gilwon Lee, *Member, IEEE*, and Brian L. Evans, *Fellow, IEEE*

**Abstract**—In this paper, we investigate hybrid analog/digital beamforming for multiple-input multiple-output (MIMO) systems with low-resolution analog-to-digital converters (ADCs) for millimeter wave (mmWave) communications. In the receiver, we propose to split the analog combining subsystem into a channel gain aggregation stage followed by a spreading stage. Both stages use phase shifters. Our goal is to design the two-stage analog combiner to optimize mutual information (MI) between the transmitted and quantized signals by effectively managing quantization error. To this end, we formulate an unconstrained MI maximization problem without a constant modulus constraint on analog combiners, and derive a two-stage analog combining solution. The solution achieves the optimal scaling law with respect to the number of radio frequency chains and maximizes the MI for homogeneous singular values of a MIMO channel. We further develop a two-stage analog combining algorithm to implement the derived solution for mmWave channels. By decoupling channel gain aggregation and spreading functions from the derived solution, the proposed algorithm implements the two functions by using array response vectors and a discrete Fourier transform matrix under the constant modulus constraint on each matrix element. Therefore, the proposed algorithm provides a near optimal solution for the unconstrained problem, whereas conventional hybrid approaches offer a near optimal solution only for a constrained problem. The closed-form approximation of the ergodic rate is derived for the algorithm, showing that a practical digital combiner with two-stage analog combining also achieves the optimal scaling law. Simulation results validate the algorithm performance and the derived ergodic rate.

**Index Terms**—Two-stage analog combining structure, low-resolution ADCs, mutual information, ergodic rate.

## I. INTRODUCTION

Millimeter wave communications have emerged as a promising technology for 5G communications [1], [2]. Utilizing multi-gigahertz bandwidth in 30-300 GHz frequency ranges enables cellular networks to achieve an order of magnitude increase in achievable rate [3], and a large number of antennas can be packed into transceivers with very small antenna spacing by leveraging the very small wavelength. Due to the large number of radio frequency (RF) chains and power-demanding high-resolution ADCs coupled with high sampling rates, however, the significant power consumption at the receivers becomes one of the primary challenges to resolve. In this paper, we consider hybrid MIMO receivers with low-resolution ADCs for mmWave communications to address such a challenge by reducing both the number of RF

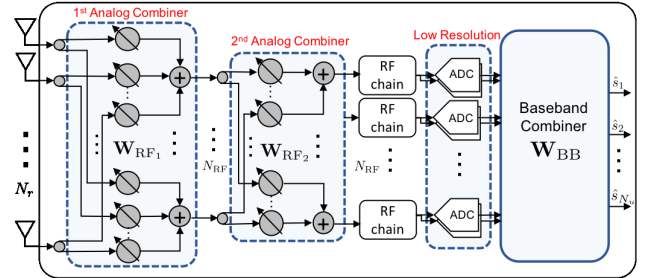


Figure 1. A receiver architecture with two-stage analog combining, low-resolution ADCs and digital combining.

chains and quantization resolution of ADCs. We propose a two-stage analog combining receiver architecture to maximize the mutual information by effectively managing quantization error as shown in Fig. 1.

### A. Prior Work

Hybrid beamforming architectures have been widely investigated to reduce the number of RF chains with minimum communication performance degradation. Singular value decomposition (SVD)-based analog combining designs were proposed [4]–[6] as the SVD transceiver maximizes the channel capacity. In [4], hybrid precoder and combiner design methods were developed by extracting the phases of the elements of the singular vectors. Considering correlated channels, the SVD of the MIMO channel covariance matrix was used for analog combiner design to maximize mutual information in [5]. The performance of hybrid precoding systems was analyzed for MIMO downlink communications [7], [8]. It was shown that hybrid beamforming systems with a small number of RF chains can achieve the performance comparable to fully digital beamforming systems. For MIMO uplink communications, the Gram-schmidt based analog combiner design algorithm was developed in [9] to orthogonalize multiuser signals.

For mmWave channels, hybrid beamforming techniques were proposed by exploiting the limited scattering of the channels [10]–[17]. Adopting array response vectors (ARVs) for analog beamformer design, orthogonal matching pursuit (OMP)-based algorithms were developed in [10]–[14]. The proposed OMP-based algorithm in [10] approximates the minimum mean squared error (MMSE) combiner with a fewer number of RF chains than the number of antennas by using ARV-based analog combiners. The OMP-based algorithm in [10] was further improved by combining OMP and local search to reduce the computational complexity [13] and by iteratively updating the phases of the phase shifters [14].

J. Choi, and B. L. Evans are with the Wireless Networking and Communication Group (WNCG), Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78701 USA. (e-mail: jinseok-choi89@utexas.edu, bevans@ece.utexas.edu). G. Lee is with Intel Corporation, Santa Clara, CA 95054 USA. (e-mail: gilwon.lee30@gmail.com). The authors at The University of Texas at Austin were supported by gift funding from Huawei Technologies.

A channel estimation technique was also proposed by using hierarchical multi-resolution codebook-based ARVs with low training overhead in [11]. By leveraging the sparse nature of mmWave channels, the proposed algorithms with ARV-based analog beamformers achieved the comparable performance with greatly reduced cost and power consumption compared to fully digital systems.

While the previous studies [4]–[17] considered infinite-resolution ADCs in hybrid MIMO systems, hybrid beamforming systems with low-resolution ADCs were investigated in [18]–[23] to take advantage of both the hybrid beamforming and low-resolution ADC architectures. The proposed algorithm in [18] attempted to design an analog combiner by minimizing the MSE including the quantization error. The analog combiner, however, is not constrained with a constant modulus, and the entire combining matrix needs to be designed for each transmitted symbol separately. Without considering the coarse quantization effect in combiner design, bit allocation techniques [19] and user scheduling methods [20] were developed for a given ARV-based analog combiner. In [21], [22], an alternating projection method was adopted to implement SVD-based analog combiners. The performance analysis of hybrid MIMO systems with low-resolution ADCs in [21] showed the superior tradeoff between performance and power consumption compared to fully digital systems and hybrid systems with infinite-resolution ADCs. In [23], a subarray antenna structure was considered, and an ARV-based combining algorithm was used to select the ARV that maximizes the aggregated channel gain. Although the analysis in [21]–[23] provided useful insights for the hybrid architecture with low-resolution ADCs such as the achievable rate and power tradeoff, the quantization error was not explicitly taken into account in the hybrid beamformer design. Consequently, considering the coarse quantization effect in the analog combiner design is still an open question.

## B. Contributions

In this paper, we derive a near optimal analog combining solution for an unconstrained MI maximization problem in hybrid MIMO systems with low-resolution ADCs. We, then, propose a two-stage analog combining architecture to properly implement the derived solution under a constant modulus constraint on each phase shifter. Splitting the solution into a channel gain aggregation stage by using ARVs and a gain spreading stage by using a discrete Fourier transform (DFT) matrix, the two-stage analog combining structure realizes the derived near optimal combining solution with phase shifter-based analog combiners for mmWave communications. The contributions of this paper can be summarized as follows:

- Without imposing a constant modulus constraint on an analog combiner, we formulate an unconstrained MI maximization problem for a hybrid MIMO system with low-resolution ADCs. For a general channel, we derive a near optimal analog combining solution which consists of (1) any semi-unitary matrix that includes the singular vectors of the signal space in the channel matrix and (2) any unitary matrix with constant modulus. The first and second parts

in the derived solution can be considered as a channel gain aggregation function that collects the entire channel gains into the lower dimension and a spreading function that reduces quantization error by spreading the aggregated gains over RF chains, respectively. We show that the derived solution achieves the optimal scaling law with respect to the number of RF chains and maximizes the MI when the singular values of a MIMO channel are the same.

- We further develop an ARV-based two-stage analog combining algorithm to implement the derived solution for mmWave channels under the constant modulus constraint on each phase shifter. Decoupling the channel gain aggregation and spreading functions from the solution, the algorithm implements the aggregation and spreading functions by using ARVs and a DFT matrix without losing the optimality of the solution in the large antenna array regime. Therefore, the two-stage analog combiner obtained from the proposed algorithm under the constant modulus constraint also provides a near optimal solution for the unconstrained MI maximization problem, whereas conventional hybrid approaches offer a near optimal solution only for a constrained problem. Since the DFT matrix is independent of channels, only passive phase shifters need to be appended to a conventional hybrid MIMO architecture with marginal complexity and cost increase, while achieving a large MI gain.
- We derive a closed-form approximation of the ergodic rate with a maximum ratio combining (MRC) digital combiner for the proposed algorithm. The derived rate characterizes the ergodic rate performance of the proposed two-stage analog combining architecture in terms of the system parameters including quantization resolution. The derived rate reveals that the ergodic rate of the MRC combiner achieves the same optimal scaling law with the proposed two-stage analog combiner by reducing the quantization error as the number of RF chains increases.

Simulation results demonstrate that the proposed two-stage analog combining algorithm outperforms conventional algorithms and validate the derived ergodic rate.

*Notation:*  $\mathbf{A}$  is a matrix and  $\mathbf{a}$  is a column vector.  $\mathbf{A}^H$  and  $\mathbf{A}^T$  denote conjugate transpose and transpose.  $[\mathbf{A}]_{i,:}$  and  $\mathbf{a}_i$  indicate the  $i$ th row and column vector of  $\mathbf{A}$ . We denote  $a_{i,j}$  or  $[\mathbf{A}]_{i,j}$  as the  $\{i, j\}$ th element of  $\mathbf{A}$  and  $a_i$  as the  $i$ th element of  $\mathbf{a}$ .  $\lambda_i\{\mathbf{A}\}$  denotes the  $i$ -th largest singular value of  $\mathbf{A}$ .  $\mathcal{CN}(\mu, \sigma^2)$  is the complex Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ .  $\mathbb{E}[\cdot]$  and  $\mathbb{V}[\cdot]$  represent an expectation and variance operators, respectively. The correlation matrix is denoted as  $\mathbf{R}_{\mathbf{xy}} = \mathbb{E}[\mathbf{xy}^H]$ . The diagonal matrix  $\text{diag}\{\mathbf{A}\}$  has  $\{a_{i,i}\}$  at its  $i$ th diagonal entry, and  $\text{diag}\{\mathbf{a}\}$  or  $\text{diag}\{\mathbf{a}^T\}$  has  $\{a_i\}$  at its  $i$ th diagonal entry.  $\text{blkdiag}\{\mathbf{A}_1, \dots, \mathbf{A}_N\}$  is a block diagonal matrix with diagonal entries  $\mathbf{A}_1, \dots, \mathbf{A}_N$ .  $\mathbf{I}$  denotes the identity matrix with a proper dimension and we indicate the dimension  $N$  by  $\mathbf{I}_N$  if necessary.  $\mathbf{0}$  denotes a matrix that has all zeros in its elements with a proper dimension.  $\|\mathbf{A}\|$  represents  $L_2$  norm.  $|\cdot|$  indicates an absolute value, cardinality, and determinant for a scalar value  $a$ , a set  $\mathcal{A}$ , and a matrix  $\mathbf{A}$ , respectively.  $\text{Tr}\{\cdot\}$  is a trace operator and  $x(N) \sim y(N)$  indicates  $\lim_{N \rightarrow \infty} \frac{x}{y} = 1$ .

## II. SYSTEM MODEL

We consider single-cell uplink wireless communications in which the BS is equipped with  $N_r$  receive antennas and  $N_{\text{RF}}$  RF chains with  $N_{\text{RF}} < N_r$ . The antennas are uniform linear arrays (ULA), and each RF chain is followed by a pair of low-resolution ADCs. We assume that the BS serves  $N_u$  users each with a single transmit antenna with  $N_u \leq N_{\text{RF}}$ .

### A. Channel Model

The channel  $\mathbf{h}_{\gamma,k}$  of user  $k$  is assumed to be the sum of the contributions of scatterers that contribute  $L_k$  propagation paths to the channel  $\mathbf{h}_{\gamma,k}$  [24]. For mmWave channels, the number of channel paths  $L_k$  is expected to be small due to the limited scattering [2]. The discrete-time narrowband channel of user  $k$  can be modeled as

$$\mathbf{h}_{\gamma,k} = \frac{1}{\sqrt{\gamma_k}} \mathbf{h}_k = \sqrt{\frac{N_r}{\gamma_k L_k}} \sum_{\ell=1}^{L_k} g_{\ell,k} \mathbf{a}(\phi_{\ell,k}) \quad (1)$$

where  $\gamma_k$  denotes the pathloss of user  $k$ ,  $g_{\ell,k}$  is the complex gain of the  $\ell$ th propagation path of user  $k$ , and  $\mathbf{a}(\phi_{\ell,k})$  is the ARV of the receive antennas corresponding to the azimuth AoA of the  $\ell$ th path of the  $k$ th user  $\phi_{\ell,k} \in [-\pi/2, \pi/2]$ . The complex channel gain  $g_{\ell,k}$  follows an independent and identically distributed (i.i.d.) complex Gaussian distribution,  $g_{\ell,k} \stackrel{i.i.d.}{\sim} \mathcal{CN}(0,1)$ . The ARV  $\mathbf{a}(\theta)$  for the ULA antennas of the BS is given as

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N_r}} \left[ 1, e^{-j\pi\vartheta}, e^{-j2\pi\vartheta}, \dots, e^{-j(N_r-1)\pi\vartheta} \right]^T$$

where the spatial angle  $\vartheta = \frac{2d}{\lambda} \sin(\theta)$  is related to the physical AoA  $\theta$ ,  $d$  is the distance between antennas, and  $\lambda$  is the signal wave length. We use  $\phi$  and  $\theta$  to denote the physical AoAs of a user channel and physical angles of analog combiners, respectively. We also use  $\varphi$  and  $\vartheta$  to denote the spatial angles for  $\phi$  and  $\theta$ , respectively, where  $\varphi, \vartheta \in [-1, 1]$ .

### B. Signal and Quantization Model

For simplicity, we consider a homogeneous long-term received SNR network<sup>1</sup> where a conventional uplink power control compensates for the pathloss and shadowing effect to achieve the same long-term received SNR target for all users in the cell [25], [26]. Let  $\mathbf{x} = \mathbf{P}\mathbf{s}$  be the transmitted user signals where  $\mathbf{P} = \text{diag}\{\sqrt{\rho\gamma_1}, \dots, \sqrt{\rho\gamma_{N_u}}\}$  is the transmit power matrix and  $\mathbf{s}$  is the  $N_u \times 1$  transmitted symbol vector from  $N_u$  users. Further, let  $\mathbf{H}_\gamma = \mathbf{H}\mathbf{B}$  represent the  $N_r \times N_u$  channel matrix where  $\mathbf{B} = \text{diag}\{\sqrt{1/\gamma_1}, \dots, \sqrt{1/\gamma_{N_u}}\}$ . The received baseband analog signal vector is given as

$$\mathbf{r} = \mathbf{H}_\gamma \mathbf{x} + \mathbf{n} = \mathbf{H}\mathbf{B}\mathbf{s} + \mathbf{n} = \sqrt{\rho} \mathbf{H}\mathbf{s} + \mathbf{n}$$

where  $\mathbf{n}$  indicates the  $N_r \times 1$  additive white noise vector. We assume zero mean and unit variance for the user symbols  $\mathbf{s}$  and noise  $\mathbf{n}$ . The noise follows the complex Gaussian distribution  $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$  and thus, we consider  $\rho$  to be the SNR.

<sup>1</sup>We remark that the derived analysis in this paper can also be applicable to a heterogeneous long-term received SNR network with minor modification.

After the BS receives the signals from users, the signals are combined via two analog combiners as shown in Fig. 1. Then, the received baseband analog signal vector becomes

$$\begin{aligned} \mathbf{y} &= \sqrt{\rho} \mathbf{W}_{\text{RF}_2}^H \mathbf{W}_{\text{RF}_1}^H \mathbf{H}\mathbf{s} + \mathbf{W}_{\text{RF}_2}^H \mathbf{W}_{\text{RF}_1}^H \mathbf{n} \\ &= \sqrt{\rho} \mathbf{W}_{\text{RF}}^H \mathbf{H}\mathbf{s} + \mathbf{W}_{\text{RF}}^H \mathbf{n} \end{aligned} \quad (2)$$

where  $\mathbf{W}_{\text{RF}} = \mathbf{W}_{\text{RF}_1} \mathbf{W}_{\text{RF}_2}$  denotes the two-stage analog combiner,  $\mathbf{W}_{\text{RF}_1} \in \mathbb{C}^{N_r \times N_{\text{RF}}}$  is the first analog combiner, and  $\mathbf{W}_{\text{RF}_2} \in \mathbb{C}^{N_r \times N_{\text{RF}}}$  is the second analog combiner. Each real and imaginary part of the combined signal (2) are quantized at ADCs with  $b$  quantization bits. Assuming a MMSE scalar quantizer and Gaussian signaling  $\mathbf{s} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_u})$ , we adopt an additive quantization noise model (AQNM) [27] which shows reasonable accuracy in the low to medium SNR ranges [28]. The AQNM approximates the quantization process in linear form, which is equivalent to the approximation with Bussgang decomposition for low-resolution ADCs [29]. The quantized signal vector is expressed as [27], [29]

$$\mathbf{y}_q = \mathcal{Q}(\mathbf{y}) = \alpha_b \sqrt{\rho} \mathbf{W}_{\text{RF}}^H \mathbf{H}\mathbf{s} + \alpha_b \mathbf{W}_{\text{RF}}^H \mathbf{n} + \mathbf{q} \quad (3)$$

where  $\mathcal{Q}(\cdot)$  is the element-wise quantizer, the scalar quantization gain is  $\alpha_b = 1 - \beta_b$  where  $\beta_b = \mathbb{E}[|y - y_q|^2] / \mathbb{E}[|y|^2]$ , and  $\mathbf{q}$  denotes the quantization noise vector. For  $b > 5$  quantization bits,  $\beta_b$  is approximated as  $\beta_b \approx \frac{\pi\sqrt{3}}{2} 2^{-2b}$ . For  $b \leq 5$ , the values of  $\beta_b$  are listed in Table 1 in [30]. The quantization noise vector  $\mathbf{q}$  is uncorrelated to the quantization input  $\mathbf{y}$  and follows the complex Gaussian distribution  $\mathbf{q} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{\mathbf{q}\mathbf{q}})$ , where the covariance matrix is given as [27]

$$\mathbf{R}_{\mathbf{q}\mathbf{q}} = \alpha_b \beta_b \text{diag}\{\rho \mathbf{W}_{\text{RF}}^H \mathbf{H}\mathbf{H}^H \mathbf{W}_{\text{RF}} + \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}}\}. \quad (4)$$

Then, a digital combiner  $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{N_{\text{RF}} \times N_{\text{RF}}}$  is applied to the quantized signal in (3) as

$$\mathbf{z} = \alpha_b \sqrt{\rho} \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{H}\mathbf{s} + \alpha_b \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{n} + \mathbf{W}_{\text{BB}}^H \mathbf{q}. \quad (5)$$

## III. OPTIMALITY OF TWO-STAGE ANALOG COMBINING

In this section, we provide a near optimal structure for the first and second analog combiners  $\mathbf{W}_{\text{RF}_1}$ ,  $\mathbf{W}_{\text{RF}_2}$  in low-resolution ADC systems for a general channel. To this end, we first formulate an unconstrained MI maximization problem without a constant modulus condition on the analog combiner  $\mathbf{W}_{\text{RF}}$ . Then, we derive a near optimal solution for the unconstrained problem, which can be splitted into two different functions corresponding to the two-stage analog combiner.

We consider the MI between the transmit symbols  $\mathbf{s}$  and quantized signals  $\mathbf{y}_q$  under the AQNM model as a measure to maximize. The MI is given as

$$\begin{aligned} \mathcal{C}(\mathbf{W}_{\text{RF}}) & \\ &= \log_2 \left| \mathbf{I}_{N_{\text{RF}}} + \rho \alpha_b^2 (\alpha_b^2 \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} + \mathbf{R}_{\mathbf{q}\mathbf{q}})^{-1} \mathbf{W}_{\text{RF}}^H \mathbf{H}\mathbf{H}^H \mathbf{W}_{\text{RF}} \right|. \end{aligned} \quad (6)$$

Using (6), we formulate the maximum MI problem by only assuming a semi-unitary constraint on the analog combiner  $\mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} = \mathbf{I}_{N_{\text{RF}}}$  as in [21] to keep the effective noise being white Gaussian noise. Accordingly, the relaxed MI maximization problem is formulated as

$$\mathcal{P}1 : \mathbf{W}_{\text{RF}}^{\text{opt}} = \underset{\mathbf{W}_{\text{RF}}}{\text{argmax}} \mathcal{C}(\mathbf{W}_{\text{RF}}), \text{ s.t. } \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} = \mathbf{I}. \quad (7)$$

Under the perfect quantization system where the number of quantization bits is assumed to be infinite, the optimal analog combiner for the problem  $\mathcal{P}1$  is given as the matrix  $\mathbf{U}_{1:N_{\text{RF}}}$  that consists of the first  $N_{\text{RF}}$  left singular vectors of  $\mathbf{H}$ . The optimal solution  $\mathbf{W}_{\text{RF}}^{\text{opt}}$  of the problem  $\mathcal{P}1$  with a finite number of quantization bits, however, is still not known. We first derive an optimal scaling law with respect to the number of RF chains  $N_{\text{RF}}$ , and provide a solution that achieves the scaling law.

**Theorem 1** (Optimal scaling law). *For fixed  $N_{\text{RF}}/N_r = \kappa$  with  $\kappa \in (0, 1)$ , the MI with the optimal combiner  $\mathbf{W}_{\text{RF}}^{\text{opt}}$  for the problem  $\mathcal{P}1$  scales with  $N_{\text{RF}}$  as*

$$\mathcal{C}(\mathbf{W}_{\text{RF}}^{\text{opt}}) \sim N_u \log_2 N_{\text{RF}} \quad (8)$$

and this optimal scaling law can be achieved by using  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF}_1}^* \mathbf{W}_{\text{RF}_2}^*$  such that:

- (i)  $\mathbf{W}_{\text{RF}_1}^* = [\mathbf{U}_{1:N_u} \mathbf{U}_\perp]$ , and
- (ii)  $\mathbf{W}_{\text{RF}_2}^*$  is any  $N_{\text{RF}} \times N_{\text{RF}}$  unitary matrix that satisfies the constant modulus condition on its elements,

where  $\mathbf{U}_{1:N_u}$  is the matrix of the left-singular vectors corresponding to the first  $N_u$  largest singular values of  $\mathbf{H}$  and  $\mathbf{U}_\perp$  denotes the matrix of any orthonormal vectors whose column space is orthogonal to that of  $\mathbf{U}_{1:N_u}$ .

*Proof.* Since the optimal solution for  $\mathcal{P}1$  is not known, we first derive an upper bound of  $\mathcal{C}(\mathbf{W}_{\text{RF}})$  and its scaling law with respect to  $N_{\text{RF}}$ . We, then, show that adopting  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF}_1}^* \mathbf{W}_{\text{RF}_2}^*$ , which satisfies the conditions (i) and (ii) in Theorem 1, achieves the same scaling law of the upper bound.

An arbitrary semi-unitary analog combiner  $\mathbf{W}_{\text{RF}}$  can be decomposed into

$$\mathbf{W}_{\text{RF}} = [\mathbf{U}_\parallel \mathbf{U}_\perp] \bar{\mathbf{W}}_{\text{RF}}, \quad (9)$$

where  $\mathbf{U}_\parallel$  is an  $N_r \times m$  matrix composed of  $m$  orthonormal basis vectors whose column space is in the subspace of  $\text{Span}(\mathbf{u}_1, \dots, \mathbf{u}_{N_u})$  with  $1 \leq m \leq N_u$ ,  $\mathbf{U}_\perp$  is an  $N_r \times (N_{\text{RF}} - m)$  matrix composed of  $(N_{\text{RF}} - m)$  orthonormal basis vectors whose column space is in the subspace of  $\text{Span}^\perp(\mathbf{u}_1, \dots, \mathbf{u}_{N_u})$ , and  $\bar{\mathbf{W}}_{\text{RF}}$  is an  $N_{\text{RF}} \times N_{\text{RF}}$  unitary matrix. Here,  $\mathbf{u}_i$  is the  $i$ -th left-singular vector of  $\mathbf{H}$ . Using (9), the term  $\mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{H}^H \mathbf{W}_{\text{RF}}$  in (6) can be re-written as

$$\begin{aligned} & \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{H}^H \mathbf{W}_{\text{RF}} \\ &= \bar{\mathbf{W}}_{\text{RF}}^H [\mathbf{U}_\parallel \mathbf{U}_\perp]^H \mathbf{U} \mathbf{U}^H [\mathbf{U}_\parallel \mathbf{U}_\perp] \bar{\mathbf{W}}_{\text{RF}} \\ &= \bar{\mathbf{W}}_{\text{RF}}^H \underbrace{\begin{bmatrix} \mathbf{U}_\parallel^H \mathbf{U}_{1:N_u} \mathbf{\Lambda}_{N_u} \mathbf{U}_{1:N_u}^H \mathbf{U}_\parallel & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\triangleq \mathbf{Q}} \bar{\mathbf{W}}_{\text{RF}} \end{aligned} \quad (10)$$

where  $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_{N_u}, 0, \dots, 0\} \in \mathbb{C}^{N_r \times N_r}$ ,  $\mathbf{\Lambda}_{N_u} = \text{diag}\{\lambda_1, \dots, \lambda_{N_u}\}$ ,  $\lambda_i$  is the  $i$ th largest singular value of  $\mathbf{H} \mathbf{H}^H$ , and  $\mathbf{U}_{1:N_r} = [\mathbf{u}_1, \dots, \mathbf{u}_{N_r}]$ . The matrix  $\mathbf{Q}$  has  $m$  ranks and can be decomposed into  $\mathbf{Q} = \mathbf{U}_\mathbf{Q} \mathbf{\Lambda} \mathbf{U}_\mathbf{Q}^H$ , where  $\mathbf{U}_\mathbf{Q}$  is the  $N_{\text{RF}} \times N_{\text{RF}}$  matrix consisting of  $N_{\text{RF}}$  singular vectors of  $\mathbf{Q}$ ; and  $\bar{\mathbf{\Lambda}} = \text{diag}\{\bar{\lambda}_1, \dots, \bar{\lambda}_m, 0, \dots, 0\} \in \mathbb{C}^{N_{\text{RF}} \times N_{\text{RF}}}$ . Here,  $\bar{\lambda}_i$  is the  $i$ th largest singular value of  $\mathbf{Q}$ . Since  $\mathbf{U}_\mathbf{Q}$  is unitary,  $\bar{\mathbf{W}}_{\text{RF}}$  can be re-expressed as

$$\bar{\mathbf{W}}_{\text{RF}} = \mathbf{U}_\mathbf{Q} \bar{\mathbf{W}}_{\text{RF}}. \quad (11)$$

and  $\bar{\mathbf{W}}_{\text{RF}}$  is still unitary. Substituting (11) into (10), we have  $\mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{H}^H \mathbf{W}_{\text{RF}} = \bar{\mathbf{W}}_{\text{RF}}^H \mathbf{\Lambda} \bar{\mathbf{W}}_{\text{RF}}$  and the MI in (6) becomes

$$\begin{aligned} & \mathcal{C}(\mathbf{W}_{\text{RF}}) \\ &= \log_2 \left| \mathbf{I} + \frac{\alpha_b}{\beta_b} \text{diag}^{-1} \left\{ \bar{\mathbf{W}}_{\text{RF}}^H \mathbf{\Lambda} \bar{\mathbf{W}}_{\text{RF}} + \frac{1}{\beta_b \rho} \mathbf{I} \right\} \bar{\mathbf{W}}_{\text{RF}}^H \mathbf{\Lambda} \bar{\mathbf{W}}_{\text{RF}} \right|. \end{aligned} \quad (12)$$

Let  $\mathbf{G} = \bar{\mathbf{W}}_{\text{RF}}^H \mathbf{\Lambda}^{1/2} = [\mathbf{G}_{\text{sub}} \mathbf{0}]$ , where  $\mathbf{G}_{\text{sub}}$  is the  $N_{\text{RF}} \times m$  submatrix of  $\mathbf{G}$ . Then, the MI can be upper bounded as

$$\begin{aligned} & \mathcal{C}(\mathbf{W}_{\text{RF}}) \\ &= \log_2 \left| \mathbf{I}_{N_{\text{RF}}} + \frac{\alpha_b}{\beta_b} \mathbf{G}^H \text{diag}^{-1} \left\{ \|\mathbf{G}\|_{i,:}^2 + \frac{1}{\beta_b \rho} \right\} \mathbf{G} \right| \\ &= \log_2 \left| \mathbf{I}_m + \frac{\alpha_b}{\beta_b} \mathbf{G}_{\text{sub}}^H \text{diag}^{-1} \left\{ \|\mathbf{G}_{\text{sub}}\|_{i,:}^2 + \frac{1}{\beta_b \rho} \right\} \mathbf{G}_{\text{sub}} \right| \\ &\stackrel{(a)}{=} \log_2 \left| \mathbf{I}_m + \frac{\alpha_b}{\beta_b} \tilde{\mathbf{G}}_{\text{sub}}^H \tilde{\mathbf{G}}_{\text{sub}} \right| \\ &= \sum_{i=1}^m \log_2 \left( 1 + \frac{\alpha_b}{\beta_b} \lambda_i \{ \tilde{\mathbf{G}}_{\text{sub}}^H \tilde{\mathbf{G}}_{\text{sub}} \} \right) \\ &\stackrel{(b)}{\leq} m \log_2 \left( 1 + \frac{\alpha_b}{\beta_b m} \sum_{i=1}^m \lambda_i \{ \tilde{\mathbf{G}}_{\text{sub}}^H \tilde{\mathbf{G}}_{\text{sub}} \} \right) \\ &\stackrel{(c)}{=} m \log_2 \left( 1 + \frac{\alpha_b}{\beta_b m} \sum_{i=1}^{N_{\text{RF}}} \frac{\|\mathbf{G}_{\text{sub}}\|_{i,:}^2}{\|\mathbf{G}_{\text{sub}}\|_{i,:}^2 + \frac{1}{\beta_b \rho}} \right) \end{aligned} \quad (13)$$

where (a) follows by letting  $\tilde{\mathbf{G}}_{\text{sub}}$  be the matrix whose each row  $i$  is given as  $i$ -th row of  $\mathbf{G}_{\text{sub}}$  normalized by  $(\|\mathbf{G}_{\text{sub}}\|_{i,:}^2 + \frac{1}{\beta_b \rho})^{1/2}$ ; (b) comes from Jensen's inequality and the concavity of  $\log_2(1+x)$  for  $x > 0$ ; and (c) is from

$$\sum_{i=1}^m \lambda_i \{ \tilde{\mathbf{G}}_{\text{sub}}^H \tilde{\mathbf{G}}_{\text{sub}} \} = \text{Tr} \{ \tilde{\mathbf{G}}_{\text{sub}}^H \tilde{\mathbf{G}}_{\text{sub}} \} = \sum_{i=1}^{N_{\text{RF}}} \frac{\|\mathbf{G}_{\text{sub}}\|_{i,:}^2}{\|\mathbf{G}_{\text{sub}}\|_{i,:}^2 + \frac{1}{\beta_b \rho}}.$$

The upper bound of  $\mathcal{C}(\mathbf{W}_{\text{RF}})$  in (13) can further be upper bounded by  $m \log_2(1 + \frac{\alpha_b N_{\text{RF}}}{\beta_b m})$  because  $\frac{\|\mathbf{G}_{\text{sub}}\|_{i,:}^2}{\|\mathbf{G}_{\text{sub}}\|_{i,:}^2 + \frac{1}{\beta_b \rho}} < 1$ . Since the derivative of this bound with respect to  $m$  is positive for  $m > 0$  with any given  $\alpha_b, N_{\text{RF}} > 0$ , it is maximized when  $m = N_u$ , and thus, it scales as  $N_u \log_2 N_{\text{RF}}$ , as  $N_{\text{RF}} \rightarrow \infty$ .

Now, we prove that the scaling law can be achieved by the two-stage analog combiner  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF}_1}^* \mathbf{W}_{\text{RF}_2}^*$  in Theorem 1. Let  $\mathbf{C} \triangleq \mathbf{W}_{\text{RF}_2}^{*H} \mathbf{\Lambda}_{N_{\text{RF}}} \mathbf{W}_{\text{RF}_2}^*$ . From  $\mathbf{W}_{\text{RF}}^{*H} \mathbf{H} \mathbf{H}^H \mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF}_2}^{*H} \mathbf{\Lambda}_{N_{\text{RF}}} \mathbf{W}_{\text{RF}_2}^* = \mathbf{C}$  where  $\mathbf{\Lambda}_{N_{\text{RF}}} = \text{diag}\{\lambda_1, \dots, \lambda_{N_u}, 0, \dots, 0\} \in \mathbb{C}^{N_{\text{RF}} \times N_{\text{RF}}}$  and (12), we have

$$\begin{aligned} & \mathcal{C}(\mathbf{W}_{\text{RF}}^*) \\ &= \log_2 \left| \mathbf{I}_{N_{\text{RF}}} + \frac{\alpha_b}{\beta_b} \text{diag}^{-1} \left\{ \mathbf{C} + \frac{1}{\beta_b \rho} \mathbf{I}_{N_{\text{RF}}} \right\} \mathbf{C} \right| \end{aligned} \quad (14)$$

$$\stackrel{(a)}{=} \log_2 \left| \mathbf{I} + \frac{\alpha_b}{\beta_b} \left( \frac{\sum_{i=1}^{N_u} \lambda_i}{N_{\text{RF}}} + \frac{1}{\beta_b \rho} \right)^{-1} \mathbf{W}_{\text{RF}_2}^{*H} \mathbf{\Lambda}_{N_{\text{RF}}} \mathbf{W}_{\text{RF}_2}^* \right| \quad (15)$$

$$\begin{aligned} &= \sum_{k=1}^{N_u} \log_2 \left( 1 + \frac{\alpha_b \rho N_{\text{RF}} \lambda_k}{N_{\text{RF}} + (1 - \alpha_b) \rho \sum_{i=1}^{N_u} \lambda_i} \right) \\ &= \sum_{k=1}^{N_u} \log_2 \left( 1 + \frac{\alpha_b \rho N_{\text{RF}} \lambda_k / N_r}{\kappa + (1 - \alpha_b) \rho \sum_{i=1}^{N_u} \lambda_i / N_r} \right) \end{aligned} \quad (16)$$

$\stackrel{(b)}{\sim} N_u \log_2 N_{\text{RF}}$ , as  $N_{\text{RF}} \rightarrow \infty$ .

Here,  $(a)$  is from that all diagonal entries of  $\mathbf{W}_{\text{RF}_2}^{*H} \mathbf{\Lambda}_{N_{\text{RF}}} \mathbf{W}_{\text{RF}_2}^*$  are the same as  $d_j = \frac{\sum_{i=1}^{N_u} \lambda_i}{N_{\text{RF}}}$ , for  $j = 1, \dots, N_{\text{RF}}$  because of the constant modulus property of  $\mathbf{W}_{\text{RF}_2}^*$ ; (b) follows from the fact that as  $N_{\text{RF}} \rightarrow \infty$ , i.e., as  $N_r \rightarrow \infty$ , we have  $\frac{1}{N_r} \mathbf{H}^H \mathbf{H} \rightarrow \text{diag}\{\frac{1}{L_1} \sum_{\ell=1}^{L_1} |g_{\ell,1}|^2, \dots, \frac{1}{L_{N_u}} \sum_{\ell=1}^{L_{N_u}} |g_{\ell,N_u}|^2\}$  [31] by the channel model in (1) without the pathloss component and the law of large numbers, which implies

$$\frac{\lambda_i}{N_r} \rightarrow \frac{1}{L_i} \sum_{\ell=1}^{L_i} |g_{\ell,i}|^2 < \infty, \text{ for } i = 1, \dots, N_u.$$

This completes the proof of Theorem 1.  $\blacksquare$

We note from (14) that  $\mathbf{W}_{\text{RF}_1}^*$  of the two-stage analog combining solution  $\mathbf{W}_{\text{RF}}^*$  aggregates all channel gains into the smaller dimension and provides  $(N_{\text{RF}} - N_u)$  extra dimensions. Then, as observed in (15),  $\mathbf{W}_{\text{RF}_2}^*$  spreads the aggregated channels gains over all  $N_{\text{RF}}$  dimensions, which reduces the quantization error by exploiting the extra dimensions. Accordingly, as the number of RF chains  $N_{\text{RF}}$  increases, the proposed solution  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF}_1}^* \mathbf{W}_{\text{RF}_2}^*$  achieves the optimal scaling law (8) by reducing the quantization error.

**Corollary 1.** *The conventional optimal solution  $\mathbf{W}_{\text{RF}}^{\text{cv}} = [\mathbf{U}_{1:N_u} \mathbf{U}_{\perp}]$  for perfect quantization systems cannot achieve the optimal scaling law (8) in coarse quantization systems, and it is upper bounded by*

$$\mathcal{C}(\mathbf{W}_{\text{RF}}^{\text{cv}}) < \mathcal{C}_{\text{svd}}^{\text{ub}} = N_u \log_2 \left( 1 + \frac{\alpha_b}{1 - \alpha_b} \right). \quad (17)$$

*Proof.* From (14), we have the following MI by setting  $\mathbf{W}_{\text{RF}_2} = \mathbf{I}$ :

$$\begin{aligned} \mathcal{C}(\mathbf{W}_{\text{RF}}^{\text{cv}}) &= \log_2 \left| \mathbf{I} + \frac{\alpha_b}{\beta_b} \text{diag}^{-1} \left\{ \mathbf{\Lambda}_{N_{\text{RF}}} + \frac{1}{\beta_b \rho} \mathbf{I} \right\} \mathbf{\Lambda}_{N_{\text{RF}}} \right| \\ &= \sum_{i=1}^{N_u} \log_2 \left( 1 + \frac{\alpha_b \lambda_i}{\beta_b \lambda_i + 1/\rho} \right) \stackrel{(a)}{<} N_u \log_2 \left( 1 + \frac{\alpha_b}{\beta_b} \right). \end{aligned}$$

where (a) comes from  $\rho > 0$ .  $\blacksquare$

Corollary 1 shows that the conventional optimal analog combiner  $\mathbf{W}_{\text{RF}}^{\text{cv}}$  can capture all channel gains but the MI does not scale as that of  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF}_1}^* \mathbf{W}_{\text{RF}_2}^*$ . Since all channel gains after processed through  $\mathbf{W}_{\text{RF}}^{\text{cv}}$  are concentrated on only  $N_u$  RF chains out of  $N_{\text{RF}}$  RF chains, using  $\mathbf{W}_{\text{RF}}^{\text{cv}}$  results in severe quantization errors at each of the  $N_u$  RF chains. Although the channel gains  $\{\lambda_i\}$  increase as  $N_r$  increases, the quantization errors also increase in proportion to the channel gains for  $\mathcal{C}(\mathbf{W}_{\text{RF}}^{\text{cv}})$ , yielding only the bounded MI in (17).

Again, unlike the conventional solution, the additional second stage analog combiner  $\mathbf{W}_{\text{RF}_2}^*$  proposed in Theorem 1 spreads the channel gains captured by the first stage combiner  $\mathbf{W}_{\text{RF}_1}^*$  to all  $N_{\text{RF}}$  RF chains evenly, leading to achieving the optimal scaling law by greatly alleviating quantization errors. Intuitively, adopting the second combiner  $\mathbf{W}_{\text{RF}_2}^*$  results in distributing the burden of ADCs confined in few RF chains over all available ADCs of the total RF chains. Later, we show that such performance gain from adopting the two-stage analog combining structure can be significant even with a reasonable number of RF chains.

**Theorem 2.** *For the case of homogeneous singular values of  $\mathbf{H}^H \mathbf{H}$  where all singular values  $\{\lambda_i\}$  are equal, the two-stage analog combining solution  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF}_1}^* \mathbf{W}_{\text{RF}_2}^*$  in Theorem 1 maximizes the MI in (7) with finite  $N_{\text{RF}}$ , i.e.,*

$$\mathbf{W}_{\text{RF}}^* = \arg \max_{\mathbf{W}_{\text{RF}}} \mathcal{C}(\mathbf{W}_{\text{RF}})$$

$$\text{s.t. } \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} = \mathbf{I}_{N_{\text{RF}}} \text{ and } \lambda_1 = \dots = \lambda_{N_u} = \lambda,$$

and the corresponding optimal MI is given as

$$\mathcal{C}_{\text{opt}} \triangleq \mathcal{C}(\mathbf{W}_{\text{RF}}^*) = N_u \log_2 \left( 1 + \frac{\alpha_b \lambda N_{\text{RF}}}{\lambda N_u (1 - \alpha_b) + N_{\text{RF}}/\rho} \right). \quad (18)$$

*Proof.* Recall  $\mathbf{G} = \overline{\mathbf{W}}_{\text{RF}}^H \bar{\mathbf{\Lambda}}^{1/2} = [\mathbf{G}_{\text{sub}} \mathbf{0}]$  in the proof of Theorem 1, where  $\mathbf{G}_{\text{sub}}$  is the  $N_{\text{RF}} \times m$  submatrix of  $\mathbf{G}$  and  $\bar{\mathbf{\Lambda}} = \text{diag}\{\bar{\lambda}_1, \dots, \bar{\lambda}_m, 0, \dots, 0\}$  is the diagonal matrix composed of the singular values of  $\mathbf{Q}$ , defined in (10). From the assumption of  $\lambda_1 = \dots = \lambda_{N_u} = \lambda$ , we have

$$\begin{aligned} \max_{\mathbf{x} \in \mathbb{C}^{N_{\text{RF}}}: \|\mathbf{x}\|=1} \mathbf{x}^H \mathbf{Q} \mathbf{x} &= \max_{\mathbf{y} \in \mathbb{C}^m: \|\mathbf{y}\|=1} \lambda \|\mathbf{U}_{1:N_u}^H \mathbf{U}_{\parallel} \mathbf{y}\|^2 \\ &\stackrel{(a)}{\leq} \max_{\mathbf{y} \in \mathbb{C}^m: \|\mathbf{y}\|=1} \lambda \|\mathbf{U}_{1:N_u}^H\|^2 \|\mathbf{U}_{\parallel}\|^2 \|\mathbf{y}\|^2 \\ &= \lambda, \end{aligned}$$

where (a) comes from the sub-multiplicativity of the norm, and the last equality holds by  $\|\mathbf{U}_{1:N_u}^H\| = 1$  and  $\|\mathbf{U}_{\parallel}\| = 1$ . This implies the singular values of  $\mathbf{Q}$  are bounded as  $\bar{\lambda}_i \leq \lambda$  for  $i = 1, \dots, m$ . Hence,  $\|[\mathbf{G}_{\text{sub}}]_{j,:}\|^2$  is maximized for any given  $\overline{\mathbf{W}}_{\text{RF}}$  when  $\bar{\lambda}_i$  achieves  $\lambda$  for all  $i = 1, \dots, m$ .

We consider the upper bound of  $\mathcal{C}(\mathbf{W}_{\text{RF}})$  in (13) and define

$$\mathbf{G}_{\text{sub}}^* = \overline{\mathbf{W}}_{\text{RF}}^H \begin{bmatrix} \sqrt{\lambda} \mathbf{I}_m \\ \mathbf{0} \end{bmatrix}.$$

Then, (13) is further upper bounded as

$$\begin{aligned} \mathcal{C}(\mathbf{W}_{\text{RF}}) &\leq m \log_2 \left( 1 + \frac{\alpha_b}{\beta_b m} \sum_{i=1}^{N_{\text{RF}}} \frac{\|[\mathbf{G}_{\text{sub}}^*]_{i,:}\|^2}{\|[\mathbf{G}_{\text{sub}}^*]_{i,:}\|^2 + \frac{1}{\beta_b \rho}} \right) \\ &\stackrel{(a)}{\leq} m \log_2 \left( 1 + \frac{\alpha_b N_{\text{RF}}}{\beta_b m} \frac{\sum_{i=1}^{N_{\text{RF}}} \|[\mathbf{G}_{\text{sub}}^*]_{i,:}\|^2}{\sum_{i=1}^{N_{\text{RF}}} \|[\mathbf{G}_{\text{sub}}^*]_{i,:}\|^2 + \frac{N_{\text{RF}}}{\beta_b \rho}} \right) \\ &\stackrel{(b)}{=} m \log_2 \left( 1 + \frac{\alpha_b \lambda N_{\text{RF}}}{\lambda m \beta_b + N_{\text{RF}}/\rho} \right), \quad (19) \end{aligned}$$

where (a) holds by Jensen's inequality and the concavity of  $\frac{x}{x+1}$  for  $x > 0$ ; and (b) comes from  $\sum_{i=1}^{N_{\text{RF}}} \|[\mathbf{G}_{\text{sub}}^*]_{i,:}\|^2 = \|\mathbf{G}_{\text{sub}}^*\|_F^2 = \lambda m$ . Note that (19) is maximized when  $m = N_u$  since the derivative of (19) with respect to  $m$  is positive for  $m > 0$  for any given  $\alpha_b, \lambda, \rho, N_{\text{RF}} > 0$ . By substituting  $\lambda_1 = \dots = \lambda_{N_u} = \lambda$  into (16), it can be shown that the upper bound of  $\mathcal{C}(\mathbf{W}_{\text{RF}})$  in (19) with  $m = N_u$  can be achieved by adopting  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF}_1}^* \mathbf{W}_{\text{RF}_2}^*$ . This completes the proof of Theorem 2.  $\blacksquare$

Theorem 2 shows the optimality of the proposed two-stage analog combining solution  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF}_1}^* \mathbf{W}_{\text{RF}_2}^*$  in maximizing the MI for any number of RF chains  $N_{\text{RF}} \geq N_u$  with homogeneous singular values. We note that such optimality of  $\mathbf{W}_{\text{RF}}^*$  can be nearly achieved for a fixed number of users in large-scale MIMO systems as shown in Remark 1.

**Remark 1.** From Theorem 2, the two-stage analog combining solution  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF1}}^* \mathbf{W}_{\text{RF2}}^*$  in Theorem 1 maximizes the MI for  $\mathcal{P}1$  as well as achieves the optimal scaling law (8) in homogeneous massive MIMO networks with a large number of antennas  $N_r$ , where each channel element  $h_{ij} \stackrel{i.i.d.}{\sim} \mathcal{CN}(0,1)$ . This is because as the number of receive antennas  $N_r$  increases,  $\frac{1}{N_r} \mathbf{H}^H \mathbf{H} \rightarrow \mathbf{I}_{N_u}$ , i.e.,  $\frac{1}{N_r} \lambda_i \rightarrow 1, \forall i$  [32].

Figure 2 shows the simulation results of the MI of the proposed two-stage analog combiner  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF1}}^* \mathbf{W}_{\text{RF2}}^*$  in Theorem 1 and the conventional analog combiner  $\mathbf{W}_{\text{RF}}^{\text{cv}}$  in Corollary 1 which is optimal for infinite-resolution ADC systems. Here, we use  $\mathbf{W}_{\text{RF1}}^* = \mathbf{W}_{\text{RF}}^{\text{cv}} = \mathbf{U}_{1:N_{\text{RF}}}$  and  $\mathbf{W}_{\text{RF2}}^* = \mathbf{W}_{\text{DFT}}$ , where  $\mathbf{W}_{\text{DFT}}$  is an  $N_{\text{RF}} \times N_{\text{RF}}$  normalized DFT matrix, and consider Rayleigh MIMO channels described in Remark 1. As shown in Fig. 2(a), the MI of the proposed two-stage analog combiner almost achieves the optimal MI  $C_{\text{opt}}$  (18) in Theorem 2 with  $\lambda/N_r = 1$  even in the regime of a finite  $N_r$ . We further note that compared with the conventional one-stage combiner  $\mathbf{W}_{\text{RF}}^{\text{cv}}$  converging to the upper limit  $C_{\text{svd}}^{\text{ub}}$ , the MI of the two-stage analog combiner logarithmically increases without a limit as  $N_r$  increases with  $\kappa \approx 1/3$ . This follows the optimal scaling law in Theorem 1.

Fig. 2(b) shows the MI simulation results with respect to the SNR  $\rho$ . The two-stage combiner  $\mathbf{W}_{\text{RF}}^* = \mathbf{W}_{\text{RF1}}^* \mathbf{W}_{\text{RF2}}^*$  yields superior MI performance to that of  $\mathbf{W}_{\text{RF}}^{\text{cv}}$ , and the MI of  $\mathbf{W}_{\text{RF}}^*$  converges to  $N_u \log_2 \left( 1 + \frac{\alpha_b N_{\text{RF}}}{(1-\alpha_b) N_u} \right)$ , which is obtained from  $C_{\text{opt}}$  (18) with  $\rho \rightarrow \infty$ . Therefore, the MI gap between the upper limits of the two combiners ( $\mathbf{W}_{\text{RF}}^*, \mathbf{W}_{\text{RF}}^{\text{cv}}$ ) is

$$\Delta = N_u \left( \log_2 \left( 1 + \frac{\alpha_b N_{\text{RF}}}{(1-\alpha_b) N_u} \right) - \log_2 \left( 1 + \frac{\alpha_b}{1-\alpha_b} \right) \right). \quad (20)$$

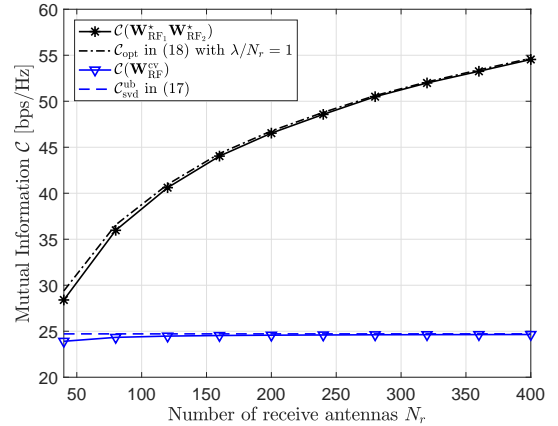
Since  $N_{\text{RF}} \geq N_u$  is considered in this paper, the proposed two-stage combiner  $\mathbf{W}_{\text{RF}}^*$  always yields the higher upper limit of the MI than the SVD-based one-stage combiner  $\mathbf{W}_{\text{RF}}^{\text{cv}}$ .

#### IV. TWO-STAGE ANALOG COMBINING ALGORITHM

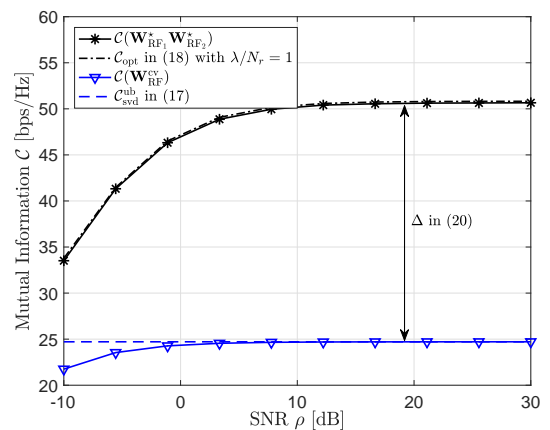
In the previous section, we derived the analog combining solution for the unconstrained problem  $\mathcal{P}1$ . However, the constant modulus constraint on each matrix element should be taken into account in designing analog combiners since it is implemented using phase shifters. We further consider a pre-defined set of phases with a finite cardinality for phase shifters. Considering channels known at the receiver, we propose a codebook-based two-stage analog combining algorithm for mmWave communications.

##### A. Proposed Two-Stage Analog Combining Algorithm

Theorem 1 provides a practical analog combiner structure that is implementable with a two-stage analog combiner  $\mathbf{W}_{\text{RF}} = \mathbf{W}_{\text{RF1}} \mathbf{W}_{\text{RF2}}$ : the first analog combiner and the second analog combiner can be considered as a channel gain aggregation matrix and spreading matrix, respectively. Leveraging such insight and the finding in the following Corollary 2, we propose an ARV-based two-stage analog combining (ARV-TSAC) algorithm for mmWave channels.



(a)  $C$  vs.  $N_r$



(b)  $C$  vs.  $\rho$

Figure 2. The simulation results of the MI with the proposed two-stage analog combining solution  $\mathbf{W}_{\text{RF1}}^* \mathbf{W}_{\text{RF2}}^*$  and the conventional optimal analog combiner  $\mathbf{W}_{\text{RF}}^{\text{cv}}$  in the Rayleigh MIMO channels: (a) for  $(\rho, N_{\text{RF}}, N_u, b) = (5 \text{ dB}, \lceil \frac{N_r}{3} \rceil, 8, 2)$  as  $N_r$  increases, and (b) for  $(N_r, N_{\text{RF}}, N_u, 2) = (256, \lceil \frac{N_r}{3} \rceil, 8, 2)$  as  $\rho$  increases.

**Corollary 2.** When the number of channel paths  $L_k$  is limited, the optimal scaling in (8) can be achieved by using  $\tilde{\mathbf{W}}_{\text{RF}}^* = \mathbf{W}_{\text{AoA}} \mathbf{W}_{\text{RF2}}^*$  as  $N_r \rightarrow \infty$  for fixed  $\kappa \in (0, 1)$ , where  $\mathbf{W}_{\text{AoA}} = [\mathbf{A}_{\text{AoA}}, \mathbf{A}_{\text{AoA}}^\perp]$ ,  $\mathbf{A}_{\text{AoA}} = [\mathbf{a}(\phi_{1,1}), \mathbf{a}(\phi_{2,1}), \dots, \mathbf{a}(\phi_{L N_u, N_u})]$ , and  $\mathbf{A}_{\text{AoA}}^\perp$  is an  $N_r \times (N_{\text{RF}} - \sum_{k=1}^{N_u} L_k)$  matrix composed of orthonormal basis vectors whose column space is in  $\text{Span}^\perp(\mathbf{A}_{\text{AoA}})$ .

*Proof.* See Appendix A. ■

According to Corollary 2, using ARVs provides a fair trade-off between practicality in implementation and performance. To design the first analog combiner  $\mathbf{W}_{\text{RF1}}$ , we adopt an ARV-codebook based maximum channel gain aggregation approach to collect most channel gains into the lower signal dimension by exploiting the sparse nature of mmWave channels. We set the codebook of the evenly spaced spatial angles  $\mathcal{V} = \{\vartheta_1, \dots, \vartheta_{|\mathcal{V}|}\}$ . Since selecting  $N_{\text{RF}}$  ARVs out of the total  $|\mathcal{V}|$  ARVs in the codebook requires  $\binom{|\mathcal{V}|}{N_{\text{RF}}}$  search complexity for the exhaustive method, we propose a greedy-based algorithm to find the best  $N_{\text{RF}}$  ARVs with greatly reduced complexity.

**Algorithm 1:** ARV-based TSAC

---

1 **Initialization:** set  $\mathbf{W}_{\text{RF}_1} = \text{empty matrix}$ ,  $\mathbf{H}_{\text{rm}} = \mathbf{H}$ ,  
and  $\mathcal{V} = \{\vartheta_1, \dots, \vartheta_{|\mathcal{V}|}\}$  where  $\vartheta_n = \frac{2n}{|\mathcal{V}|} - 1$

2 **for**  $i = 1 : N_{\text{RF}}$  **do**

3     Maximum channel gain aggregation  
       (a)  $\mathbf{a}(\vartheta^*) = \text{argmax}_{\vartheta \in \mathcal{V}} \|\mathbf{a}(\vartheta)^H \mathbf{H}_{\text{rm}}\|^2$   
       (b)  $\mathbf{W}_{\text{RF}_1} = [\mathbf{W}_{\text{RF}_1} \mid \mathbf{a}(\vartheta^*)]$   
       (c)  $\mathbf{H}_{\text{rm}} = \mathcal{P}_{\mathbf{a}(\vartheta^*)}^\perp \mathbf{H}_{\text{rm}}$ , where  $\mathcal{P}_{\mathbf{a}(\vartheta^*)}^\perp = \mathbf{I} - \mathbf{a}(\vartheta^*)\mathbf{a}(\vartheta^*)^H$   
       (d)  $\mathcal{V} = \mathcal{V} \setminus \{\vartheta^*\}$

4 **end**

5 Set  $\mathbf{W}_{\text{RF}_2} = \mathbf{W}_{\text{DFT}}$  where  $\mathbf{W}_{\text{DFT}}$  is a normalized  
 $N_{\text{RF}} \times N_{\text{RF}}$  DFT matrix.

6 **return**  $\mathbf{W}_{\text{RF}_1}$  and  $\mathbf{W}_{\text{RF}_2}$ ;

---

Algorithm 1 describes the proposed ARV-TSAC method. In Step (a), the ARV with the spatial angle  $\vartheta^*$  which captures the largest channel gain in the remaining channel dimensions  $\mathbf{H}_{\text{rm}}$  is selected and it composes a column of the first analog combiner in Step (b). In Step (c), the channel matrix on the remaining dimensions  $\mathbf{H}_{\text{rm}}$  is projected onto the subspace of  $\text{Span}^\perp(\mathbf{a}(\vartheta^*))$  to remove the channel gain on the space of the selected ARV. Algorithm 1 repeats these steps until  $N_{\text{RF}}$  ARVs are selected from the codebook  $\mathcal{V}$ .

**Remark 2.** We can implement the second-stage analog combiner that satisfies the condition (ii) of Theorem 1 by adopting a normalized  $N_{\text{RF}} \times N_{\text{RF}}$  DFT matrix, i.e.,  $\mathbf{W}_{\text{RF}_2}^* = \mathbf{W}_{\text{DFT}}$ .

Employing the DFT matrix for the second analog combiner  $\mathbf{W}_{\text{RF}_2} = \mathbf{W}_{\text{DFT}}$  (or any unitary matrix with constant modulus) offers benefits in reducing implementation complexity and power consumption since  $\mathbf{W}_{\text{DFT}}$  does not depend on the channel  $\mathbf{H}$  and can be constructed by using passive (or fixed) analog phase shifters. Accordingly, although the additional  $N_{\text{RF}}^2$  fully-connected passive phase shifters for the second analog combiner add to the complexity of the proposed architecture in physical area and power consumption, it can be implemented with very low complexity and power consumption in the practical system. Furthermore, if  $N_{\text{RF}}$  is a power of two, the fast Fourier transform version of the DFT calculation can be implemented, which reduces the number of additional passive phase shifters to  $N_{\text{RF}} \log_2 N_{\text{RF}}$ .

### B. Performance Analysis

In this subsection, we analyze the ergodic sum rate of the ARV-TSAC algorithm with an MRC baseband combiner. Once we derive the closed-form ergodic rate, we compare the rate with the one without the second analog combiner  $\mathbf{W}_{\text{RF}_2}$  to quantify the ergodic rate gain from employing  $\mathbf{W}_{\text{RF}_2}$ . To this end, we adopt a virtual channel representation [33] for analytic tractability which captures the sparse property of mmWave channels [17], [34]. Under the virtual channel representation, the channel vector  $\mathbf{h}_k$  in (1) can be modeled as

$$\mathbf{h}_k = \sqrt{\frac{N_r}{L_k}} \mathbf{A} \tilde{\mathbf{g}}_k = \tilde{\mathbf{A}} \tilde{\mathbf{h}}_{b,k}$$

where  $\tilde{\mathbf{h}}_{b,k} = \sqrt{\frac{N_r}{L_k}} \tilde{\mathbf{g}}_k$  is the  $L_k$ -sparse beamspace channel of user  $k$ , i.e.,  $\tilde{\mathbf{g}}_k$  has  $L_k$  nonzero entries  $\overset{i.i.d.}{\sim} \mathcal{CN}(0, 1)$ , and  $\mathbf{A} = [\mathbf{a}(\varphi_1), \dots, \mathbf{a}(\varphi_{N_r})]$  with uniformly spaced spatial angles  $\varphi_i$ .

Under this representation, we consider the case where the codebook size of Algorithm 1 is equal to the number of antennas  $|\mathcal{V}| = N_r$ . Accordingly, the first analog combiner is the  $N_r \times N_{\text{RF}}$  submatrix of  $\mathbf{A}$  which captures the most channel gain,  $\mathbf{W}_{\text{RF}_1} = \mathbf{A}_{\text{sub}}$ . We assume that  $\mathbf{W}_{\text{RF}_1}$  captures all channel propagation paths from  $N_u$  users [19], [35], i.e.,  $L_k$  channels paths for each user fall within  $N_{\text{RF}}$  RF chains. For simplicity, we further assume  $L_k = L$ ,  $\forall k$ , in the analysis<sup>2</sup>. Thus, after combining with  $\mathbf{W}_{\text{RF}_1} = \mathbf{A}_{\text{sub}}$ , the channel becomes  $\mathbf{H}_b = \mathbf{W}_{\text{RF}_1}^H \mathbf{H}$ , and the channel vector of user  $k$  with the reduced dimension  $\mathbf{h}_{b,k} \in \mathbb{C}^{N_{\text{RF}}}$  is

$$\mathbf{h}_{b,k} = \sqrt{\frac{N_r}{L}} \mathbf{g}_k. \quad (21)$$

We consider  $L$  nonzero channel gains to be uniformly distributed within each user channel  $\mathbf{h}_{b,k}$  and use an indicator function  $\mathbb{1}_{\{i \in \mathcal{A}\}}$  to characterize the channel sparsity where  $\mathbb{1}_{\{i \in \mathcal{A}\}} = 1$  if  $i \in \mathcal{A}$ , and  $\mathbb{1}_{\{i \in \mathcal{A}\}} = 0$  otherwise. Utilizing  $\mathbb{1}_{\{\cdot\}}$ , we model the  $\ell$ th complex path gain of user  $k$  as

$$g_{\ell,k} = \xi_{\ell,k} \mathbb{1}_{\{\ell \in \mathcal{P}_k\}}, \quad \ell = 1, \dots, N_{\text{RF}}, \quad k = 1, \dots, N_u$$

where  $\xi_{\ell,k} \overset{i.i.d.}{\sim} \mathcal{CN}(0, 1)$ ,  $\forall \ell, k$  and  $\mathcal{P}_k = \{i \mid g_{i,k} \neq 0, i = 1, \dots, N_{\text{RF}}\}$  is the nonzero index set.

We consider the MRC combiner  $\mathbf{W}_{\text{BB}} = \bar{\mathbf{H}}_b$  where  $\bar{\mathbf{H}}_b = \mathbf{W}_{\text{RF}_2}^H \mathbf{W}_{\text{RF}_1}^H \mathbf{H}$ , and the received signal  $k$  in (5) becomes

$$z_k = \alpha_b \sqrt{\rho} \bar{\mathbf{h}}_{b,k}^H \bar{\mathbf{h}}_{b,k} s_k + \alpha_b \sqrt{\rho} \sum_{i \neq k} \bar{\mathbf{h}}_{b,k}^H \bar{\mathbf{h}}_{b,i} s_i + \alpha_b \bar{\mathbf{h}}_{b,k}^H \mathbf{W}_{\text{RF}}^H \mathbf{n} + \bar{\mathbf{h}}_{b,k}^H \mathbf{q}. \quad (22)$$

From (22), the achievable rate of the proposed system for the MRC combiner with simplification is given as

$$r_k^{\text{mrc}} = \log_2 \left( 1 + \frac{\rho \alpha_b \|\bar{\mathbf{h}}_{b,k}\|^4}{\rho \alpha_b \sum_{i \neq k} \bar{\mathbf{h}}_{b,k}^H \bar{\mathbf{h}}_{b,i} + \|\bar{\mathbf{h}}_{b,k}\|^2 + \rho \beta_b \Psi_k} \right) \quad (23)$$

where  $\Psi_k = \bar{\mathbf{h}}_{b,k}^H \text{diag}\{\bar{\mathbf{H}}_b \bar{\mathbf{H}}_b^H\} \bar{\mathbf{h}}_{b,k}$ , and the ergodic rate is

$$\begin{aligned} \bar{r}_k^{\text{mrc}} &= \mathbb{E} \left[ r_k^{\text{mrc}} \right] \\ &= \mathbb{E} \left[ \log_2 \left( 1 + \frac{\rho \alpha_b \|\bar{\mathbf{h}}_{b,k}\|^4}{\rho \alpha_b \sum_{i \neq k} \bar{\mathbf{h}}_{b,k}^H \bar{\mathbf{h}}_{b,i} + \|\bar{\mathbf{h}}_{b,k}\|^2 + \rho \beta_b \Psi_k} \right) \right]. \end{aligned} \quad (24)$$

Since  $\mathbf{W}_{\text{RF}_2} = \mathbf{W}_{\text{DFT}}$  is unitary, we have  $\|\bar{\mathbf{h}}_{b,i}^H \bar{\mathbf{h}}_{b,j}\| = \|\mathbf{h}_{b,i}^H \mathbf{h}_{b,j}\|$ ,  $\forall i, j$ . We approximate the ergodic rate (24) as

$$\begin{aligned} \bar{r}_k^{\text{mrc}} &= \mathbb{E} \left[ \log_2 \left( 1 + \frac{\rho \alpha_b \|\mathbf{h}_{b,k}\|^4}{\rho \alpha_b \sum_{i \neq k} |\mathbf{h}_{b,k}^H \mathbf{h}_{b,i}|^2 + \|\mathbf{h}_{b,k}\|^2 + \rho \beta_b \Psi_k} \right) \right] \\ &\overset{(a)}{\approx} \log_2 \left( 1 + \frac{\rho \alpha_b \mathbb{E}[\|\mathbf{h}_{b,k}\|^4]}{\rho \alpha_b \sum_{i \neq k} \mathbb{E}[|\mathbf{h}_{b,k}^H \mathbf{h}_{b,i}|^2] + \mathbb{E}[\|\mathbf{h}_{b,k}\|^2] + \rho \beta_b \mathbb{E}[\Psi_k]} \right) \end{aligned} \quad (25)$$

<sup>2</sup>The similar results can be derived with minor changes for general  $L_k$ .

where (a) follows from Lemma 1 in [36].

We first analyze the average quantization error with two-stage analog combining and MRC  $\mathbb{E}[\Psi_k]$  in (25). Noting that  $\Psi_k = \mathbf{h}_{b,k}^H \mathbf{W}_{\text{DFT}} \text{diag}\{\mathbf{W}_{\text{DFT}}^H \mathbf{H}_b \mathbf{H}_b^H \mathbf{W}_{\text{DFT}}\} \mathbf{W}_{\text{DFT}}^H \mathbf{h}_{b,k}$ , we decompose  $\mathbb{E}[\Psi_k]$  as  $\mathbb{E}[\Psi_k] = \mathbb{E}[\Psi_k^{\text{auto}}] + \mathbb{E}[\Psi_k^{\text{cross}}]$ , and define the auto quantization noise and cross quantization noise variances as

$$\mathbb{E}[\Psi_k^{\text{auto}}] \quad (26)$$

$$= \mathbb{E}\left[\mathbf{h}_{b,k}^H \mathbf{W}_{\text{DFT}} \text{diag}\{\mathbf{W}_{\text{DFT}}^H \mathbf{h}_{b,k} \mathbf{h}_{b,k}^H \mathbf{W}_{\text{DFT}}\} \mathbf{W}_{\text{DFT}}^H \mathbf{h}_{b,k}\right],$$

$$\mathbb{E}[\Psi_k^{\text{cross}}] \quad (27)$$

$$= \mathbb{E}\left[\mathbf{h}_{b,k}^H \mathbf{W}_{\text{DFT}} \text{diag}\{\mathbf{W}_{\text{DFT}}^H \mathbf{H}_{b \setminus k} \mathbf{H}_{b \setminus k}^H \mathbf{W}_{\text{DFT}}\} \mathbf{W}_{\text{DFT}}^H \mathbf{h}_{b,k}\right]$$

where  $\mathbf{H}_{b \setminus k}$  denotes the channel matrix  $\mathbf{H}_b$  without its  $k$ th column. Then, (26) and (27) represent the average quantization errors for the associated user caused by the associated user itself and other users, respectively.

**Lemma 1.** *For the considered mmWave channel, the auto quantization noise variance for the two-stage analog combining of the ARV-TSAC algorithm with MRC (26) is derived as*

$$\mathbb{E}[\Psi_k^{\text{auto}}] = \frac{2N_r^2}{N_{\text{RF}}}. \quad (28)$$

*Proof.* See Appendix B. ■

Note that the quantization noise variance decreases as the number of RF chains  $N_{\text{RF}}$  increases, which corresponds to the intuition: the second DFT analog combiner spreads the quantization noise over the  $N_{\text{RF}}$  chains and thus reduces the quantization error more as  $N_{\text{RF}}$  increases.

**Lemma 2.** *For the considered mmWave channel, the cross quantization noise variance for the two-stage analog combining of the ARV-TSAC algorithm with MRC (27) is derived as*

$$\mathbb{E}[\Psi_k^{\text{cross}}] = \frac{N_r^2(N_u - 1)}{N_{\text{RF}}}. \quad (29)$$

*Proof.* See Appendix C. ■

Since both  $\mathbb{E}[\Psi_k^{\text{auto}}]$  and  $\mathbb{E}[\Psi_k^{\text{cross}}]$  decrease with  $N_{\text{RF}}$ , the quantization error with the proposed two-stage analog combining and MRC combining is expected to decrease as  $N_{\text{RF}}$  increases, leading the ergodic rate to the same scaling law as in (8). We derive the approximated ergodic sum rate of (23) in closed form and validate the insight.

**Theorem 3.** *For the considered mmWave channel with low-resolution ADCs, the ergodic sum rate of the ARV-based TSAC method with MRC is approximated as*

$$\bar{\mathcal{R}}^{\text{mrc}} \approx N_u \log_2 \left( 1 + \frac{\rho \alpha_b N_r N_{\text{RF}} (1 + 1/L)}{N_{\text{RF}} + \rho N_r (N_u - 1) + 2\rho(1 - \alpha_b) N_r} \right). \quad (30)$$

*Proof.* See Appendix D. ■

Note that the derived ergodic rate in (30) is a function of system parameters and provides insights how the ergodic rate is improved with the proposed two-stage analog combining.

**Remark 3.** *Let  $\kappa = N_{\text{RF}}/N_r$  where  $\kappa \in (0, 1)$  is a constant value. Then, (30) can reduce to*

$$\bar{\mathcal{R}}^{\text{mrc}} \approx N_u \log_2 \left( 1 + \frac{\rho \alpha_b N_{\text{RF}} (1 + 1/L)}{\kappa + \rho(N_u - 1) + 2\rho(1 - \alpha_b)} \right). \quad (31)$$

*The ergodic sum rate in (31) achieves the optimal scaling law  $\sim N_u \log N_{\text{RF}}$  with respect to  $N_{\text{RF}}$  as in (8).*

Remark 3 shows that the optimal scaling law can be achieved by the proposed two-stage analog combining algorithm even with the practical baseband combiner. This result verifies that the two-stage analog combining architecture is effective to enhance the achievable rate in mmWave hybrid MIMO systems with low-resolution ADCs. To specify the effect of employing the second analog combiner  $\mathbf{W}_{\text{RF}_2}$ , we also derive the ergodic rate (24) without using  $\mathbf{W}_{\text{RF}_2}$ .

**Corollary 3.** *For the considered mmWave channel with low-resolution ADCs, the MRC ergodic rate of the ARV-TSAC without the second analog combiner is approximated as*

$$\begin{aligned} \bar{\mathcal{R}}_{\text{one}}^{\text{mrc}} \\ \approx N_u \log_2 \left( 1 + \frac{\rho \alpha_b N_r N_{\text{RF}} (1 + 1/L)}{N_{\text{RF}} + \rho N_r (N_u - 1) + 2\rho(1 - \alpha_b) N_r N_{\text{RF}}/L} \right). \end{aligned} \quad (32)$$

*Proof.* See Appendix E. ■

Unlike the quantization noise term  $2\rho(1 - \alpha_b)N_r$  in (30), that  $2\rho(1 - \alpha_b)N_r N_{\text{RF}}/L$  in (32) includes  $N_{\text{RF}}/L$ , which prevents the optimal scaling of the ergodic sum rate as in (8) with respect to  $N_{\text{RF}}$  for fixed  $L$ .

**Remark 4.** *Let  $\kappa = N_{\text{RF}}/N_r$  where  $\kappa \in (0, 1)$  is a constant value. Then, (32) can reduce to*

$$\bar{\mathcal{R}}_{\text{one}}^{\text{mrc}} \approx N_u \log_2 \left( 1 + \frac{\rho \alpha_b N_{\text{RF}} (1 + 1/L)}{\kappa + \rho(N_u - 1) + 2\rho(1 - \alpha_b) N_{\text{RF}}/L} \right). \quad (33)$$

*Note that unlike the ergodic rate of the two-stage analog combining  $\bar{\mathcal{R}}^{\text{mrc}}$  in (31), that of the one-stage analog combining  $\bar{\mathcal{R}}_{\text{one}}^{\text{mrc}}$  in (33) cannot achieve the optimal scaling law with respect to the number of RF chains  $N_{\text{RF}}$ .*

## V. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed two-stage analog combining algorithm in the MI and ergodic sum rate. In the simulations, we set the codebook size to be  $|\mathcal{V}| = N_r$ , which guarantees  $\mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} = \mathbf{I}_{N_{\text{RF}}}$ . Consequently, analog combiners used in the simulations are semi-unitary. To provide a reference performance of a conventional one-stage analog combining approach, we simulate a greedy-based MI maximization method which solves the following problem for the given ARV codebook in a greedy way:

$$\mathcal{P}2: \quad \mathbf{W}_{\text{RF}}^{\text{opt},c} = \underset{\mathbf{W}_{\text{RF}}}{\text{argmax}} \mathcal{C}(\mathbf{W}_{\text{RF}})$$

$$\text{s.t.} \quad \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} = \mathbf{I}, \quad |[\mathbf{W}_{\text{RF}}]_{i,j}| = \frac{1}{\sqrt{N_r}}, \forall i, j.$$

At each iteration, the greedy method searches for a single ARV from the codebook  $\mathcal{V}$  which maximizes the MI with



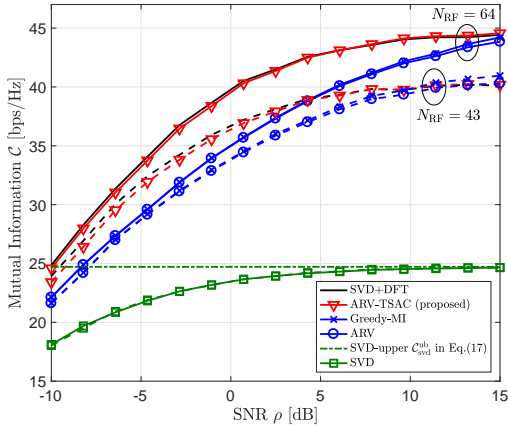


Figure 3. The MI simulation results for  $N_r = 128$  receive antennas,  $N_u = 8$  users,  $\lambda_L = 3$  average channel paths,  $b = 2$  quantization bits, and  $N_{\text{RF}} \in \{43, 64\}$  RF chains that are  $\lceil N_r/3 \rceil$  and  $\lceil N_r/2 \rceil$ , respectively.

the previously selected ARVs and thus can nearly provide the optimal MI performance of the one-stage analog combining for the given codebook.

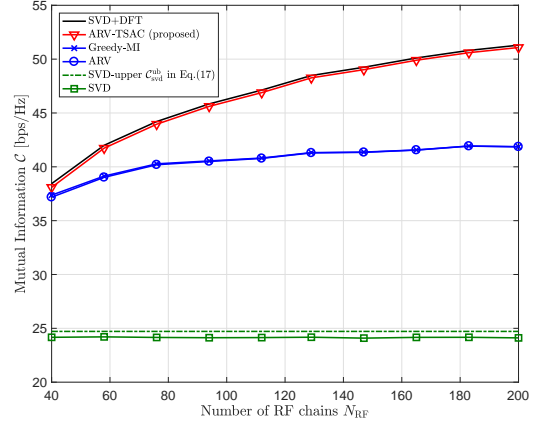
In the simulations, we evaluate the following cases:

- 1) ARV-TSAC: proposed two-stage analog combining.
- 2) ARV: one-stage analog combining with  $\mathbf{W}_{\text{RF}} = \mathbf{W}_{\text{RF}1}$  selected from the ARV-TSAC.
- 3) SVD+DFT: two-stage analog combining with  $\mathbf{W}_{\text{RF}1} = \mathbf{U}_{1:N_{\text{RF}}}$  and  $\mathbf{W}_{\text{RF}2} = \mathbf{W}_{\text{DFT}}$  based on Theorem 1.
- 4) SVD: one-stage analog combining  $\mathbf{W}_{\text{RF}} = \mathbf{U}_{1:N_{\text{RF}}}$ .
- 5) Greedy-MI: one-stage analog combining with greedy-based MI maximization.

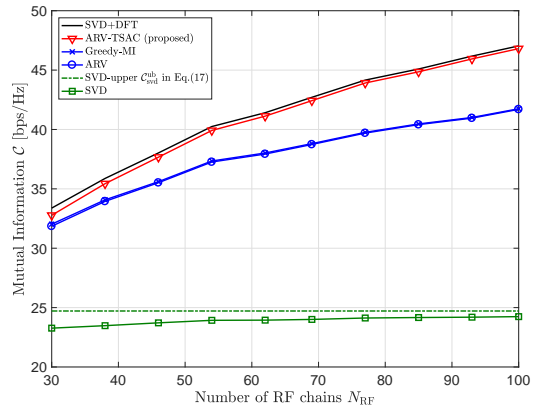
The SVD+DFT and SVD cases are infeasible in practice due to violating the constant modulus constraint, and SVD+DFT provides a tight upper bound on MI for a homogeneous singular value case from Theorem 2. Here, we adopt  $L_k = \max\{1, \text{Poisson}(\lambda_L)\}$  [37] unless mentioned otherwise, where  $\lambda_L$  is considered as the average number of channel paths.

#### A. Mutual Information

Fig. 3 shows the MI simulation results for  $N_r = 128$ ,  $N_{\text{RF}} \in \{43, 64\}$ ,  $N_u = 8$ ,  $\lambda_L = 3$ , and  $b = 2$  with respect to the SNR  $\rho$ . The proposed ARV-TSAC algorithm achieves a similar MI as does the SVD+DFT case, and they show the best MI over the most SNR values. The Greedy-MI and ARV cases provide similar MI to each other but show the MI gap from the ARV-TSAC. The gap decreases as  $\rho$  increases in the high SNR regime, and the Greedy-MI and ARV cases with  $N_{\text{RF}} = 43$  show the higher MI than SVD+DFT and ARV-TSAC in the very high SNR regime. Such phenomenon occurs as the channel environment does not guarantee the optimality condition for the two-stage analog combining solution in Theorem 2. As more RF chains are used, however, the MI gap between ARV-TSAC/SVD+DFT and Greedy-MI/ARV becomes larger and the performance reversal would happen in even the higher SNR regime. This is because the proposed two-stage analog combining can exploit more RF chains to further reduce quantization errors. The SVD



(a)  $N_r = 256$



(b)  $\kappa = 1/3$

Figure 4. The MI simulation results with  $N_u = 8$  users,  $\lambda_L = 4$  average channel paths,  $b = 2$  quantization bits, and  $\rho = 0$  dB SNR for (a)  $N_r = 256$  receive antennas and (b)  $\kappa = N_{\text{RF}}/N_r = 1/3$ .

case results in the worst MI performance and it converges to the theoretic upper bound  $C_{\text{svd}}^{\text{ub}}$  due to the quantization error.

Fig. 4 shows the MI simulation results with  $N_u = 8$ ,  $\lambda_L = 4$ ,  $b = 2$ , and  $\rho = 0$  dB in terms of  $N_{\text{RF}}$ . In Fig. 4(a),  $N_r$  is fixed to be  $N_r = 256$ . The two-stage combining cases, i.e., SVD+DFT and ARV-TSAC, show that the MI increases logarithmically with  $N_{\text{RF}}$ , and this corresponds to the scaling law derived in Theorem 1. The one-stage combining cases such as the Greedy-MI, ARV, and SVD cases, however, show a marginal increase of the MI as  $N_{\text{RF}}$  increases. In Fig. 4(b), the ratio between  $N_r$  and  $N_{\text{RF}}$  is fixed to be  $\kappa = 1/3$ . Here, the Greedy-MI and ARV cases also increase more slowly compared to the SVD+DFT and ARV-TSAC cases. This is because more channel gains can be collected as  $N_r$  increases for all cases, but the two-stage combining can reduce more quantization error as  $N_{\text{RF}}$  increases. Accordingly, the MI gap between the two-stage combining and one-stage combining cases increases as  $N_{\text{RF}}$  increases.

#### B. Ergodic Sum Rate

Now, we evaluate the ergodic rate for linear digital combiners  $\mathbf{W}_{\text{BB}}$  such as MRC, zero-forcing (ZF), and MMSE. Let  $\mathbf{H}_{\text{eq}} = \mathbf{W}_{\text{RF}}^H \mathbf{H}$ . The MRC, ZF, and MMSE combiners

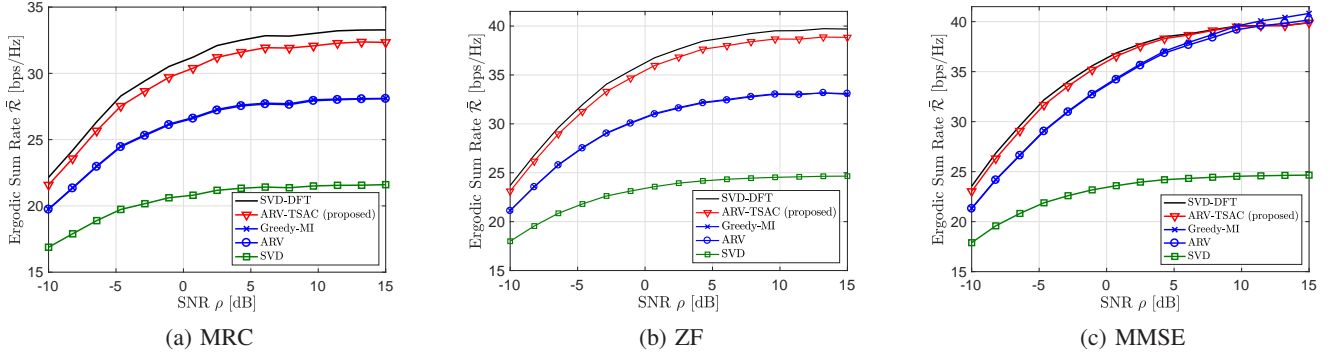


Figure 5. Simulation results of the ergodic sum rate with  $N_r = 128$  receive antennas,  $N_{\text{RF}} = 43$  RF chains,  $N_u = 8$  users,  $\lambda_L = 3$  average channel paths, and  $b = 2$  quantization bits for (a) maximum ratio combining (MRC), (b) zero-forcing (ZF), and (c) minimum mean squared error (MMSE) digital combiners.

are given as:  $\mathbf{W}_{\text{BB},\text{mrc}} = \mathbf{H}_{\text{eq}}$ ,  $\mathbf{W}_{\text{BB},\text{zf}} = \mathbf{H}_{\text{eq}}(\mathbf{H}_{\text{eq}}^H \mathbf{H}_{\text{eq}})^{-1}$ , and  $\mathbf{W}_{\text{BB},\text{mmse}} = \mathbf{R}_{\mathbf{y}_q \mathbf{y}_q}^{-1} \mathbf{R}_{\mathbf{y}_q \mathbf{x}}$ , where  $\mathbf{R}_{\mathbf{y}_q \mathbf{x}} = \alpha \rho \mathbf{H}_{\text{eq}}$  and  $\mathbf{R}_{\mathbf{y}_q \mathbf{y}_q} = \alpha^2 \rho \mathbf{H}_{\text{eq}} \mathbf{H}_{\text{eq}}^H + \alpha^2 \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} + \mathbf{R}_{\text{qq}}$ . For the given analog and digital combiners ( $\mathbf{W}_{\text{RF}}$ ,  $\mathbf{W}_{\text{BB}}$ ) with  $\mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} = \mathbf{I}_{N_{\text{RF}}}$ , the ergodic rate of user  $k$  is expressed as

$$\bar{r}_k(\mathbf{W}_{\text{RF}}, \mathbf{W}_{\text{BB}}) = \mathbb{E} \left[ \log_2 \left( 1 + \alpha_b^2 \rho |\mathbf{w}_{\text{BB},k}^H \mathbf{h}_{\text{eq},k}|^2 / \eta_{\text{BB},k} \right) \right]$$

where  $\eta_{\text{BB},k} = \alpha_b^2 \rho \sum_{u \neq k} |\mathbf{w}_{\text{BB},k}^H \mathbf{h}_{\text{eq},u}|^2 + \alpha_b^2 \|\mathbf{w}_{\text{BB},k}\|^2 + \mathbf{w}_{\text{BB},k}^H \mathbf{R}_{\text{qq}} \mathbf{w}_{\text{BB},k}$ .

Fig. 5 illustrates the ergodic sum rates with  $N_r = 128$ ,  $N_{\text{RF}} = 43$ ,  $N_u = 8$ ,  $\lambda_L = 3$ , and  $b = 2$  versus the SNR  $\rho$  for different digital combiners: (a) MRC, (b) ZF, and (c) MMSE. Similarly to the MI results, ARV-TSAC shows the comparable ergodic rate to that of SVD+DFT and outperforms the one-stage combining such as the Greedy-MI and ARV cases in most cases. We note that the SVD case also shows the worst sum rate performance in the considered systems. The gaps between the two-stage combining cases and one-stage combining cases for the MRC and ZF combiners are much larger than the gap for the MMSE combiner. In addition, SVD+DFT and ARV-TSAC with the ZF combiner achieve the ergodic rates comparable to the MMSE combiner, while the Greedy-MI and ARV cases with the ZF combiner show much lower ergodic sum rates than that with the MMSE combiner. Since the MRC and ZF combiners ignore the AWGN and quantization noise whereas the MMSE combiner does not, using the MMSE combiner improves the ergodic rate of the one-stage analog combining cases. The two-stage analog combining cases, however, already reduced the quantization noise by using the second analog combiner, and thus, they provide the MMSE-like ergodic rate performance with the ZF combiner. Therefore, the proposed two-stage analog combining with the ARV-TSAC algorithm can achieve significant rate improvement with the MRC or ZF combiners compared to the one-stage analog combining approach.

Fig. 6 provides the simulation results of the ergodic rate with the MRC digital combiner for  $N_u = 8$ ,  $\lambda_L = 3$ , and  $\rho = 0$  dB in terms of the number of (a) RF chains  $N_{\text{RF}}$  and (b) quantization bits  $b$ . In Fig. 6(a), we consider  $b = 2$  and  $\kappa = N_{\text{RF}}/N_r = 1/3$ . The ergodic rates of SVD+DFT and ARV-TSAC are similar and both increase logarithmically with  $N_{\text{RF}}$ , whereas the ergodic rates of the Greedy-MI and ARV

cases increase more slowly. Such scaling results correspond to Remark 3 and 4. As  $N_r$  increases with a fixed  $\kappa$ , SVD+DFT and ARV-TSAC effectively reduce the more quantization error while obtaining larger channel gains, but the Greedy-MI and ARV cases only obtain larger channel gains without mitigating the quantization error. In Fig. 6(b), we consider  $N_r = 128$  and  $N_{\text{RF}} = 43$ . We note that in the low-resolution ADC regime, the ARV-TSAC algorithm achieves the ergodic rate comparable to that of SVD+DFT and shows a noticeable improvement compared to the Greedy-MI, ARV, and SVD cases. As  $b$  increases, the ergodic rates of the ARV-TSAC, Greedy-MI, and ARV algorithms converge to each other with a small gap from the SVD+DFT case. The ergodic rate of the SVD case, however, converges to that of SVD+DFT without any gap because the SVD combining is optimal in maximizing the MI of infinite-resolution ADC systems. The simulation results validate the effectiveness of the proposed two-stage combining in low-resolution ADC systems.

Finally, we validate the derived ergodic rates in Theorem 3 and Corollary 3. We consider  $N_r = 128$  receive antennas,  $N_{\text{RF}} = 43$  RF chains,  $N_u = 8$  users each with  $L = 8$  channel paths for the virtual channels, and  $b = 2$  quantization bits. In Fig. 7, the theoretical ergodic rates tightly align with the simulation results in the medium to high SNR regime, and show similar trend as the simulation results do. Thus, the derived ergodic rates can characterize the ergodic rate performance of the proposed algorithm for the two-stage analog combining system in terms of the system parameters including quantization resolution.

Overall, the two-stage analog combining structure with the ARV-TSAC algorithm almost achieves the performance of SVD+DFT that is a near optimal solution for the unconstrained problem  $\mathcal{P}1$ , while the greedy-MI and ARV algorithms provide a near optimal solution only for the constrained problem  $\mathcal{P}2$ . Since  $\mathcal{P}1$  has a larger feasible set than  $\mathcal{P}2$  to find an optimal solution for the same objective function, this leads to  $\mathcal{C}(\mathbf{W}_{\text{RF}}^{\text{opt}}) \geq \mathcal{C}(\mathbf{W}_{\text{RF}}^{\text{opt},c})$ . In this regard, the ARV-TSAC algorithm achieves the higher performance than that of the Greedy-MI and ARV algorithms in most cases. This shows that the proposed two-stage analog combining architecture with the ARV-TSAC is a practical solution suitable for the mmWave hybrid MIMO systems with low-resolution ADCs.

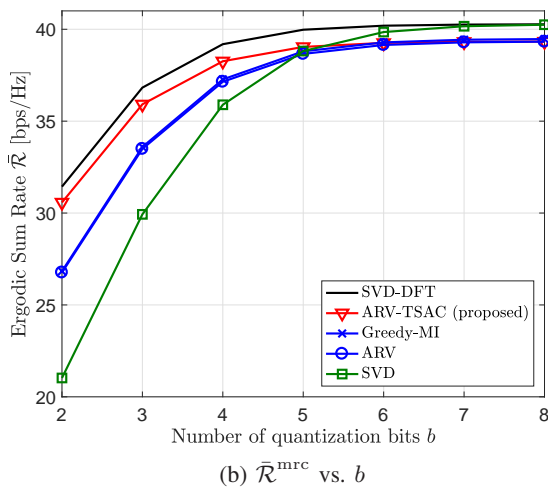
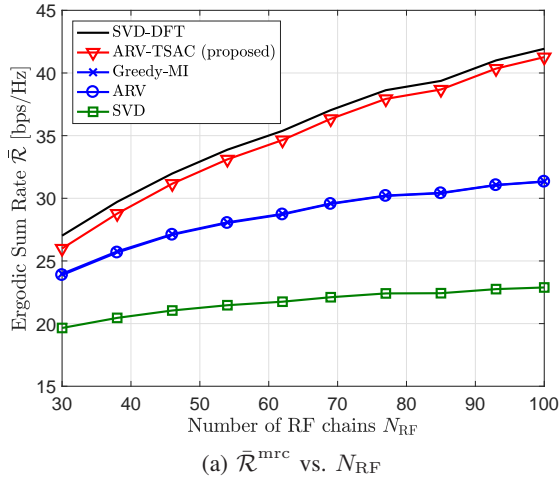


Figure 6. Simulation results of the ergodic sum rate of the MRC combiner  $\bar{\mathcal{R}}^{\text{mrc}}$  with  $N_u = 8$  users,  $\lambda_L = 3$  average channel paths, and  $\rho = 0$  dB SNR for (a)  $b = 2$  quantization bits and  $\kappa = N_{\text{RF}}/N_r = 1/3$  and (b)  $N_r = 128$  receive antennas and  $N_{\text{RF}} = 43$  RF chains.

## VI. CONCLUSION

In this paper, we derived a near optimal analog combining solution for an unconstrained MI maximization problem in hybrid MIMO systems with low-resolution ADCs. We showed optimalities of the solution in the scaling law and in maximizing the mutual information for a homogeneous channel singular value case. To implement the derived solution, we proposed a two-stage analog combining architecture that decouples the channel gain aggregation and spreading functions in the solution into two cascaded analog combiners. Accordingly, the proposed two-stage analog combining also provides a near optimal solution for the unconstrained problem whereas conventional hybrid algorithms offer a near optimal solution only for the constrained problem. In addition, we derived a closed-form approximation to the ergodic rate, which reveals that our two-stage analog combiner achieves the optimal scaling law with a practical digital combiner. Simulation results validated the key insights obtained in this paper and the derived ergodic rate, and also demonstrated that the proposed two-stage analog combining algorithm outperforms conventional algorithms.

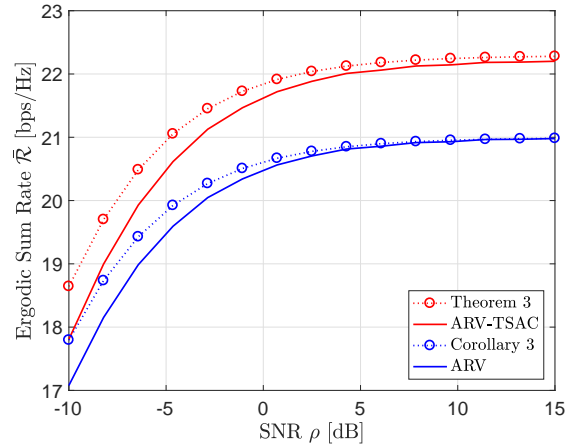


Figure 7. Comparison of the ergodic rate for the theoretical and simulation results with  $N_r = 128$  receive antennas,  $N_{\text{RF}} = 43$  RF chains,  $N_u = 8$  users each with  $L = 8$  channel paths for the virtual channels.

## APPENDIX A PROOF OF COROLLARY 2

Let  $\mathbf{H}$  be decomposed into  $\mathbf{H} = \mathbf{A}_{\text{AoA}}\mathbf{H}_V$ , where  $\mathbf{H}_V = \text{blkdiag}\{\tilde{\mathbf{g}}_1, \dots, \tilde{\mathbf{g}}_{N_u}\}$  and  $\tilde{\mathbf{g}}_k = \sqrt{\frac{N_r}{L_k}}[g_{1,k}, \dots, g_{L_k,k}]^T$ . Then, it can be shown [31] that as  $N_r \rightarrow \infty$ ,

$$\mathbf{W}_{\text{AoA}}^H \mathbf{W}_{\text{AoA}} \rightarrow \mathbf{I}_{N_{\text{RF}}}, \quad \frac{1}{\sqrt{N_r}} \mathbf{W}_{\text{AoA}}^H \mathbf{H} \rightarrow \frac{1}{\sqrt{N_r}} \begin{bmatrix} \mathbf{H}_V \\ \mathbf{0} \end{bmatrix}. \quad (34)$$

Let  $\tilde{\mathbf{H}}_V = [\mathbf{H}_V^T, \mathbf{0}^T]^T$  and  $\mathbf{C}_{\text{AoA}} = \mathbf{W}_{\text{RF}_2}^* \tilde{\mathbf{H}}_V \tilde{\mathbf{H}}_V^H \mathbf{W}_{\text{RF}_2}^*$ . Using (34), we show  $\mathcal{C}(\tilde{\mathbf{W}}_{\text{RF}})$  in (12) with  $\mathbf{W}_{\text{RF}} = \tilde{\mathbf{W}}_{\text{RF}}^*$  converges as  $N_r \rightarrow \infty$  to

$$\left( \mathcal{C}(\tilde{\mathbf{W}}_{\text{RF}}^*) - \log_2 \left[ \mathbf{I} + \frac{\alpha_b \rho}{\beta_b} \text{diag}^{-1} \left\{ \mathbf{C}_{\text{AoA}} + \frac{1}{\beta_b \rho} \mathbf{I} \right\} \mathbf{C}_{\text{AoA}} \right] \right) \rightarrow 0. \quad (35)$$

Note that each diagonal of  $\mathbf{W}_{\text{RF}_2}^* \tilde{\mathbf{H}}_V \tilde{\mathbf{H}}_V^H \mathbf{W}_{\text{RF}_2}^*$  cannot exceed  $\frac{1}{\kappa} \sum_{k=1}^{N_u} \frac{1}{L_k} (\sum_{\ell=1}^{L_k} |g_{\ell,k}|)^2 = c_1 < \infty$ . Let  $\mathcal{C}_\infty(\tilde{\mathbf{W}}_{\text{RF}}^*)$  denote the second term in (35). Then,  $\mathcal{C}_\infty(\tilde{\mathbf{W}}_{\text{RF}}^*)$  can be lower bounded as

$$\begin{aligned} \mathcal{C}_\infty(\tilde{\mathbf{W}}_{\text{RF}}^*) &> \log_2 \left| \mathbf{I}_{N_{\text{RF}}} + \frac{\alpha_b \rho}{c_1 \beta_b \rho + 1} \mathbf{W}_{\text{RF}_2}^* \tilde{\mathbf{H}}_V \tilde{\mathbf{H}}_V^H \mathbf{W}_{\text{RF}_2}^* \right| \\ &\stackrel{(a)}{\approx} N_u \log_2 N_{\text{RF}}, \quad \text{as } N_{\text{RF}} \rightarrow \infty, \end{aligned} \quad (36)$$

where (a) follows from the same reason of (b) below (16). This implies that  $\mathcal{C}(\tilde{\mathbf{W}}_{\text{RF}}^*)$  follows the optimal scaling law. ■

## APPENDIX B PROOF OF LEMMA 1

The auto quantization noise variance term in (26) can be expressed as

$$\begin{aligned} \mathbb{E}[\Psi_k^{\text{auto}}] &= \mathbb{E} \left[ \sum_{i=1}^{N_{\text{RF}}} |\mathbf{h}_{b,k}^H \mathbf{w}_i|^4 \right] = \left( \frac{N_r}{L} \right)^2 \sum_{i=1}^{N_{\text{RF}}} \mathbb{E} \left[ |\mathbf{g}_k^H \mathbf{w}_i|^4 \right] \\ &= \left( \frac{N_r}{L} \right)^2 \sum_{i=1}^{N_{\text{RF}}} \left( \mathbb{V} \left[ |\mathbf{g}_k^H \mathbf{w}_i|^2 \right] + \left( \mathbb{E} \left[ |\mathbf{g}_k^H \mathbf{w}_i|^2 \right] \right)^2 \right) \end{aligned} \quad (37)$$

where  $\mathbf{w}_i$  is the  $i$ th column of  $\mathbf{W}_{\text{DFIT}}$ . The expectation term  $\mathbb{E}[|\mathbf{g}_k^H \mathbf{w}_i|^2]$  in (37) is computed as

$$\mathbb{E}[|\mathbf{g}_k^H \mathbf{w}_i|^2] = \frac{1}{N_{\text{RF}}} \mathbb{E}\left[\sum_{\ell=1}^{N_{\text{RF}}} |g_{\ell,k}|^2\right] = \frac{L}{N_{\text{RF}}}. \quad (38)$$

Now, let  $\hat{\mathbf{w}}_i = \sqrt{N_{\text{RF}}} \mathbf{w}_i$ . Then, we can compute the variance term  $\mathbb{V}[|\mathbf{g}_k^H \mathbf{w}_i|^2]$  in (37) as

$$\begin{aligned} \mathbb{V}[|\mathbf{g}_k^H \mathbf{w}_i|^2] &= \frac{1}{N_{\text{RF}}^2} \mathbb{V}\left[\sum_{\ell=1}^{N_{\text{RF}}} |g_{\ell,k}|^2 + \sum_{\ell_1 \neq \ell_2}^{N_{\text{RF}}} g_{\ell_1,k}^* g_{\ell_2,k} \hat{w}_{\ell_1,i}^* \hat{w}_{\ell_2,i}\right] \\ &\stackrel{(a)}{=} \frac{1}{N_{\text{RF}}^2} \left( \mathbb{V}\left[\sum_{\ell=1}^{N_{\text{RF}}} |g_{\ell,k}|^2\right] + \mathbb{V}\left[\sum_{\ell_1 \neq \ell_2}^{N_{\text{RF}}} g_{\ell_1,k}^* g_{\ell_2,k} \hat{w}_{\ell_1,i}^* \hat{w}_{\ell_2,i}\right] \right) \\ &\stackrel{(b)}{=} \frac{1}{N_{\text{RF}}^2} \left( \mathbb{V}[\|\mathbf{g}_k\|^2] + \sum_{\ell_1 \neq \ell_2}^{N_{\text{RF}}} \mathbb{V}[g_{\ell_1,k}^* g_{\ell_2,k}] \right) \end{aligned} \quad (39)$$

where (a) and (b) hold as the associated terms are uncorrelated, which can be shown from straight forward mathematics, and  $|\hat{w}_{\ell,i}| = 1, \forall \ell, i$ . Since  $\|\mathbf{g}_k\|^2 \sim \chi_{2L}^2$ , which is a chi-square distribution with  $2L$  degrees of freedom, we have  $\mathbb{V}[\|\mathbf{g}_k\|^2] = L$ , and  $\mathbb{V}[g_{\ell_1,k}^* g_{\ell_2,k}]$  is computed as

$$\begin{aligned} \mathbb{V}[g_{\ell_1,k}^* g_{\ell_2,k}] &= \mathbb{V}\left[\xi_{\ell_1,k}^* \xi_{\ell_2,k} \mathbf{1}_{\{\ell_1 \in \mathcal{P}_k\}} \mathbf{1}_{\{\ell_2 \in \mathcal{P}_k\}}\right] \\ &\stackrel{(a)}{=} \mathbb{E}[|\xi_{\ell_1,k}^* \xi_{\ell_2,k}|^2] \mathbb{E}[\mathbf{1}_{\{\ell_1, \ell_2 \in \mathcal{P}_k\}}] \\ &\quad - \left(\mathbb{E}[\xi_{\ell_1,k}^* \xi_{\ell_2,k}]\right)^2 \left(\mathbb{E}[\mathbf{1}_{\{\ell_1, \ell_2 \in \mathcal{P}_k\}}]\right)^2 \\ &= \frac{L(L-1)}{N_{\text{RF}}(N_{\text{RF}}-1)}, \end{aligned}$$

where (a) holds by  $\mathbb{V}[XY] = \mathbb{E}[X^2]\mathbb{E}[Y^2] - (\mathbb{E}[X])^2(\mathbb{E}[Y])^2$  for independent  $X$  and  $Y$ . Therefore, (39) is derived as

$$\mathbb{V}[|\mathbf{g}_k^H \mathbf{w}_i|^2] = \frac{1}{N_{\text{RF}}^2} \left( L + \sum_{\ell_1 \neq \ell_2}^{N_{\text{RF}}} \frac{L(L-1)}{N_{\text{RF}}(N_{\text{RF}}-1)} \right) = \left( \frac{L}{N_{\text{RF}}} \right)^2. \quad (40)$$

Putting (38) and (40) into (37), the auto quantization noise variance  $\mathbb{E}[\Psi_k^{\text{auto}}]$  becomes (28). ■

#### APPENDIX C PROOF OF LEMMA 2

We derive the cross quantization noise variance in (27) as

$$\begin{aligned} \mathbb{E}[\Psi_k^{\text{cross}}] &= \mathbb{E}\left[\sum_{i=1}^{N_{\text{RF}}} \sum_{u \neq 1}^{N_u} \mathbf{h}_{b,k}^H \mathbf{w}_i \mathbf{w}_i^H \mathbf{h}_{b,u} \mathbf{h}_{b,u}^H \mathbf{w}_i \mathbf{w}_i^H \mathbf{h}_{b,k}\right] \\ &= \left(\frac{N_r}{L}\right)^2 \mathbb{E}_{\mathbf{g}_k} \left[ \sum_{i=1}^{N_{\text{RF}}} \sum_{u \neq 1}^{N_u} \mathbf{g}_k^H \mathbf{w}_i \mathbf{w}_i^H \mathbb{E}_{\mathbf{g}_u} [\mathbf{g}_u \mathbf{g}_u^H] \mathbf{w}_i \mathbf{w}_i^H \mathbf{g}_k \right] \\ &= \frac{N_r^2 (N_u - 1)}{L N_{\text{RF}}} \sum_{i=1}^{N_{\text{RF}}} \mathbb{E}_{\mathbf{g}_k} [\mathbf{g}_k^H \mathbf{w}_i \mathbf{w}_i^H \mathbf{g}_k] \\ &\stackrel{(a)}{=} \frac{N_r^2 (N_u - 1)}{N_{\text{RF}}} \end{aligned}$$

where (a) follows from  $\mathbb{E}[|\mathbf{g}_k^H \mathbf{w}_i|^2] = \frac{L}{N_{\text{RF}}}$  in (38). ■

#### APPENDIX D PROOF OF THEOREM 3

To compute (25), we first derive  $\mathbb{E}[\|\mathbf{h}_{b,k}\|^2]$  as

$$\mathbb{E}[\|\mathbf{h}_{b,k}\|^2] = \frac{N_r}{L} \mathbb{E}[\|\mathbf{g}_k\|^2] \stackrel{(a)}{=} N_r \quad (41)$$

where (a) follows from  $\|\mathbf{g}_k\|^2 \sim \chi_{2L}^2$ . Next, we compute  $\mathbb{E}[\|\mathbf{h}_{b,k}\|^4]$  as

$$\begin{aligned} \mathbb{E}[\|\mathbf{h}_{b,k}\|^4] &= \mathbb{V}[\|\mathbf{h}_{b,k}\|^2] + (\mathbb{E}[\|\mathbf{h}_{b,k}\|^2])^2 \\ &= \left(\frac{N_r}{L}\right)^2 \left(\mathbb{V}[\|\mathbf{g}_k\|^2] + (\mathbb{E}[\|\mathbf{g}_k\|^2])^2\right) \\ &= \frac{N_r^2(1+L)}{L}. \end{aligned} \quad (42)$$

The inter-user interference term  $\mathbb{E}[\|\mathbf{h}_{b,k}^H \mathbf{h}_{b,i}\|^2]$  is computed as

$$\begin{aligned} \mathbb{E}[\|\mathbf{h}_{b,k}^H \mathbf{h}_{b,i}\|^2] &= \left(\frac{N_r}{L}\right)^2 \mathbb{E}[|\mathbf{g}_k^H \mathbf{g}_i|^2] = \left(\frac{N_r}{L}\right)^2 \sum_{\ell=1}^{N_{\text{RF}}} \mathbb{E}[|g_{\ell,k}^* g_{\ell,i}|^2] \\ &= \left(\frac{N_r}{L}\right)^2 \sum_{\ell=1}^{N_{\text{RF}}} \mathbb{E}\left[|\xi_{\ell,k}^* \mathbf{1}_{\{\ell \in \mathcal{P}_k\}} \xi_{\ell,i} \mathbf{1}_{\{\ell \in \mathcal{P}_i\}}|^2\right] \\ &= \frac{N_r^2}{N_{\text{RF}}}. \end{aligned} \quad (43)$$

Finally, we compute the quantization variance term  $\mathbb{E}[\Psi_k]$  as

$$\begin{aligned} \mathbb{E}[\Psi_k] &= \mathbb{E}[\Psi_k^{\text{auto}}] + \mathbb{E}[\Psi_k^{\text{cross}}] \\ &\stackrel{(a)}{=} \frac{2N_r^2}{N_{\text{RF}}} + \frac{N_r^2(N_u - 1)}{N_{\text{RF}}}, \end{aligned} \quad (44)$$

where  $\mathbb{E}[\Psi_k^{\text{auto}}]$  and  $\mathbb{E}[\Psi_k^{\text{cross}}]$  are in (26) and (27), respectively, and (a) follows from Lemma 1 and Lemma 2.

Putting (41), (42), (43), and (44) into (25), we derive the approximated ergodic rate of (25) in closed form. The ergodic rate is equivalent to  $N_u$  users, which leads to the ergodic sum rate in (30). This completes the proof of Theorem 3. ■

#### APPENDIX E PROOF OF COROLLARY 3

Without the second analog combiner  $\mathbf{W}_{\text{RF}}$ , the approximated ergodic rate of user  $k$  can be computed as (25) by substituting the average quantization noise variance for the two-stage analog combining  $\mathbb{E}[\Psi_k]$  with the following average quantization noise variance:

$$\begin{aligned} \mathbb{E}[\hat{\Psi}_k] &= \mathbb{E}\left[\mathbf{h}_{b,k}^H \text{diag}\{\mathbf{H}_b \mathbf{H}_b^H\} \mathbf{h}_{b,k}\right] \\ &= \mathbb{E}\left[\left(\frac{N_r}{L}\right)^2 \sum_{\ell=1}^{N_{\text{RF}}} |g_{\ell,k}|^2 \sum_{u=1}^{N_u} |g_{\ell,u}|^2\right] \\ &= \left(\frac{N_r}{L}\right)^2 \left(\sum_{\ell=1}^{N_{\text{RF}}} \mathbb{E}[|g_{\ell,k}|^4] + \sum_{\ell=1}^{N_{\text{RF}}} \sum_{u \neq k}^{N_u} \mathbb{E}[|g_{\ell,k}|^2 |g_{\ell,u}|^2]\right). \end{aligned} \quad (45)$$

Here,  $\mathbb{E}[|g_{\ell,k}|^4]$  in (45) is computed as

$$\begin{aligned}\mathbb{E}[|g_{\ell,k}|^4] &= \mathbb{E}\left[\mathbb{1}_{\{\ell \in \mathcal{P}_k\}}\right] \mathbb{E}\left[|\xi_{\ell,k}|^4\right] \\ &= \frac{L}{N_{\text{RF}}}\left(\mathbb{V}\left[|\xi_{\ell,k}|^2\right] + \left(\mathbb{E}\left[|\xi_{\ell,k}|^2\right]\right)^2\right) \\ &= \frac{2L}{N_{\text{RF}}},\end{aligned}\quad (46)$$

and the second expectation term  $\mathbb{E}[|g_{\ell,k}|^2|g_{\ell,u}|^2]$  is derived as

$$\begin{aligned}\mathbb{E}\left[|g_{\ell,k}|^2|g_{\ell,u}|^2\right] &= \mathbb{E}\left[\mathbb{1}_{\{\ell \in \mathcal{P}_k\}}\mathbb{1}_{\{\ell \in \mathcal{P}_u\}}\right] \mathbb{E}\left[|\xi_{\ell,k}|^2|\xi_{\ell,u}|^2\right] \\ &= \left(\frac{L}{N_{\text{RF}}}\right)^2.\end{aligned}\quad (47)$$

Putting (46) and (47) into (45), we derive the average quantization noise variance for the one-stage analog combining as

$$\mathbb{E}[\hat{\Psi}_k] = N_r^2 \left( \frac{2}{L} + \frac{N_u - 1}{N_{\text{RF}}} \right).$$

This completes the proof of Corollary 3.  $\blacksquare$

#### REFERENCES

- [1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
- [2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [4] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, 2005.
- [5] P. Sudarshan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, 2006.
- [6] F. Gholam, J. Via, and I. Santamaría, "Beamforming design for simplified analog antenna combining architectures," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2373–2378, 2011.
- [7] D. Ying, F. W. Vook, T. A. Thomas, and D. J. Love, "Hybrid structure in massive MIMO: Achieving large sum rate with fewer RF chains," in *IEEE Int. Conf. Commun.*, 2015, pp. 2344–2349.
- [8] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *arXiv preprint arXiv:1601.06814*, 2016.
- [9] J. Li, L. Xiao, X. Xu, and S. Zhou, "Robust and low complexity hybrid beamforming for uplink multiuser mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1140–1143, 2016.
- [10] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.
- [11] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.
- [12] T. E. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," *IEEE Global Commun. Conf.*, 2014.
- [13] C. Rusu, R. Méndez-Rial, N. González-Prelcicy, and R. W. Heath, "Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems," in *IEEE Int. Conf. Commun.*, 2015, pp. 1340–1345.
- [14] C.-E. Chen, "An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 285–288, 2015.
- [15] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, 2014.
- [16] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, 2015.
- [17] R. Méndez-Rial, C. Rusu, N. González-Prelcicy, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.
- [18] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, 2010.
- [19] J. Choi, B. L. Evans, and A. Gatherer, "Resolution-adaptive hybrid MIMO architectures for millimeter wave communications," *IEEE Trans. Signal Process.*, vol. 65, no. 23, pp. 6201–6216, 2017.
- [20] J. Choi, G. Lee, and B. L. Evans, "User Scheduling for Millimeter Wave Hybrid Beamforming Systems with Low-Resolution ADCs," *arXiv preprint arXiv:1804.03079*, 2018.
- [21] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, 2017.
- [22] W. B. Abbas, F. Gomez-Cuba, and M. Zorzi, "Millimeter wave receiver efficiency: A comprehensive comparison of beamforming schemes with low resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8131–8146, 2017.
- [23] K. Roth, H. Pirzadeh, A. L. Swindlehurst, and J. A. Nossek, "A Comparison of Hybrid Beamforming and Digital Beamforming with Low-Resolution ADCs for Multiple Users and Imperfect CSI," *IEEE Journal Sel. Topics Signal Process.*, 2018.
- [24] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Personal Commun.*, vol. 5, no. 1, pp. 10–22, 1998.
- [25] A. Simonsson and A. Furuskar, "Uplink power control in LTE-overview and performance, subtitle: principles and benefits of utilizing rather than compensating for SINR variations," in *IEEE Veh. Technol. Conf.*, 2008, pp. 1–5.
- [26] E. Tejaswi and B. Suresh, "Survey of power control schemes for LTE uplink," *Int. Journal Computer Science and Inform. Technol.*, vol. 10, p. 2, 2013.
- [27] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Robust predictive quantization: Analysis and design via convex optimization," *IEEE Journal Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 618–632, 2007.
- [28] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *IEEE Inform. Theory and App. Work.*, Feb. 2015, pp. 191–198.
- [29] A. Mezghani and J. A. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *IEEE Int. Symp. Inform. Theory*, 2012.
- [30] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186–2189, 2015.
- [31] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *European Signal Process. Conf.*, 2014, pp. 76–80.
- [32] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590 – 3600, Nov. 2010.
- [33] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Nov. 2002.
- [34] R. W. Heath, N. Gonzalez-Prelcicy, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Feb. 2016.
- [35] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *IEEE Int. Work. Signal Process. Advances in Wireless Comm.*, 2015, pp. 146–150.
- [36] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE Journal Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, 2014.
- [37] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.