Hybrid Beamforming for Large Antenna Arrays with Phase Shifter Selection

Sohail Payami, Mir Ghoraishi, and Mehrdad Dianati
Institute for Communication Systems (ICS),
5G Innovation Centre (former CCSR), University of Surrey, UK
Email: {s.payami, m.ghoraishi, and m.dianati}@surrey.ac.uk

ABSTRACT

This paper proposes an asymptotically optimal hybrid beamforming solution for large antenna arrays by exploiting the properties of the singular vectors of the channel matrix. It is shown that the elements of the channel matrix with Rayleigh fading follow a normal distribution when large antenna arrays are employed. The proposed beamforming algorithm is effective in both sparse and rich propagation environments, and is applicable for both point-to-point and multiuser scenarios. In addition, a closed-form expression and a lower-bound for the achievable rates are derived when analog and digital phase shifters are employed. It is shown that the performance of the hybrid beamformers using phase shifters with more than 2-bits resolution is comparable with analog phase shifting. A novel phase shifter selection scheme that reduces the power consumption at the phase shifter network is proposed when the wireless channel is modeled by Rayleigh fading. Using this selection scheme, the spectral efficiency can be increased as the power consumption in the phase shifter network reduces. Compared to the scenario that all of the phase shifters are in operation, the simulation results indicate that the spectral efficiency increases when up to 50% of phase shifters are turned off.

Index Terms

Hybrid beamforming, large MIMO systems, phase shifter selection.

I. Introduction

Multiple-input-multiple-output (MIMO) techniques such as beamforming, precoding and combining can significantly improve the reliability of the transmission and increase the achievable data rates in wireless communication systems. As the number of the antenna elements at the transmitter/receiver increases, higher diversity and multiplexing gains are observed and the channel matrix tends to have favorable conditions [1]. Hence, MIMO systems with large number of antennas have attracted a lot of attention.

Depending on the structure of the antenna array, analog, digital or hybrid beamformers can be implemented. The analog approach cannot provide multiplexing gains as the antenna array is connected to the transceiver by only a single RF chain [2]. On the other hand, digital beamformers with a dedicated RF chain per antenna element can use all the degrees of freedom of the channel to transmit multiple symbols simultaneously. However, digital beamforming for large antenna arrays is not suitable for practical applications due to the system complexity, cost and power consumption [3]. In order to provide a tradeoff between performance and cost, hybrid beamformers have been proposed where a small number of RF chains are connected to a large number of antennas through a network of phase shifters [4]–[15]. This type of beamformers show a promising performance even with limited channel state information (CSI) [4]–[7]. In order to design hybrid beamformers, however, it is required to solve a complex nonconvex optimization problem due to the constant modulus constraint imposed by the phase shifters [3]. In addition, the phase shifters in practical systems have discrete resolution which converts the optimization to a computationally expensive combinatorial problem [8], [9].

In the hybrid beamforming approach, it has been shown that the baseband precoder and the RF beamformer can be designed either jointly [8]–[10] or in two stages [11]–[15]. For a point-to-point system, a joint design approach based on matching pursuit was proposed when the channel is sparse [8], [9]. In this method, firstly the singular vectors of the channel should be calculated, and then the hybrid beamformer is derived by solving an optimization problem to minimize the Euclidean distance between the matrices containing the singular vectors and the weights of the hybrid beamformer. Considering that the calculation of the singular vectors is computationally expensive, the second round of computations can cause sever delays in practical systems. In addition, the achievable spectral efficiency based on [8], [9] significantly

depends on the number of RF chains in the system and multipath components in the channel. A close to optimal performance for both rich and sparse scattering channels was proposed based on approximating the nonconvex optimization with a convex problem and using an iterative algorithm [10]. The problem associated with such iterative algorithms is that the convergence time depends on the initial conditions. Hence, the processing time to calculate the parameters of the hybrid beamformer can become a prohibitive factor in real-time systems. In the two stage design approach, the RF beamformer is calculated based on the channel matrix. Then, the baseband precoder takes the impact of the channel matrix and the RF beamformer into account. The optimal hybrid beamformer for a single stream transmission was proposed in [11]. Then, based on the simulation results it was shown that hybrid and digital beamformers can achieve a similar spectral efficiency when multiple symbols are transmitted. In this case, the optimality of the hybrid beamformer and its performance closed-form expressions were not derived. Another two stage algorithm that can achieve a close to optimal performance was reported in [13]–[15] where the RF beamformer was iteratively calculated. In [8]–[15], it is not possible to derive the closed-form expression of the performance as computer simulations are necessary to evaluate the performance. Furthermore, due to the computational delays associated with the derivation of the hybrid beamforming weights, the algorithms may not be suitable for practical systems depending on the application. In addition, the power consumption in the RF beamformer will be significantly high as each phase shifter requires some power to operate and hybrid beamformers employ a massive number of these components.

In order to address the aforementioned challenges and facilitate the implementation of hybrid beamformers, two main objectives are followed in this paper. Firstly, an asymptotically, also called virtually, optimal hybrid beamforming scheme and the closed-form expressions of the spectral efficiency for both point-to-point and multiuser scenarios in rich and sparse scattering channels are derived. Secondly, a novel phase shifter selection scheme is proposed to simultaneously increase the spectral efficiency and reduce the power consumption in the phase shifter network when rich scattering channels are considered. It is assumed that the rich and sparse scattering channels follow Rayleigh fading and geometry based models, perfect CSI is available at the transmitter and the number of the antennas are large. All the proposed schemes and the closed-form expressions in this paper are derived based on the properties of the singular vectors of the channel matrix. Using the basic characteristics of such vectors, an alternative approach

to the solution in [11] is presented. It is shown that the performance of the digital beamformers is achievable when the number of the RF chains is two times larger than the number of the transmitted symbols. In order to calculate the hybrid beamformer for the Rayleigh fading scenario when the number of the RF chains and symbols are equal, the distribution of the elements of the singular vectors of the large channel matrix are derived which, to the best knowledge of the authors, has not been previously reported. Based on this distribution, the virtually optimal hybrid beamforming schemes for both the point-to-point and multiuser scenarios are derived. Additionally, the closed-form expressions of the spectral efficiencies achieved by the proposed hybrid beamformers are calculated. It is shown that in the solution with optimum performance, the phase shifters in the RF beamformer should be set according to the phase of the elements of the singular vectors of the channel matrix when the number of the antennas are large. The advantages of the proposed approach over the stat-of-the-art is its simplicity, low computational delays and asymptotically optimal behavior. When digital phase shifters are used, a simple but effective hybrid beamforming scheme is proposed and its performance lower-bound is derived. Analytical and simulation results demonstrate that the performance of the proposed scheme with phase shifters with more than 2-bits of resolution is similar to the hybrid beamformer with analog phase shifters. Finally, a novel phase shifter selection scheme and the closed-form expression of its performance bound are presented when the channel matrix follows Rayleigh fading. The advantages of this method are two fold as the power consumption in the RF beamformer network can be reduced and the spectral efficiency can be improved at the same time. Simulation results indicate that the spectral efficiency will increase when up to 50% of the phase shifters are switched off.

This paper is organized as following, the system model and problem statement of the point-to-point system are described in sections II and III. In section IV, the hybrid beamforming scheme with analog and digital phase shifters are proposed and analyzed. The multiuser scenario and phase shifter selection are investigated in sections V and VI. Finally, the simulation results, conclusion and future works are presented in sections VII and VIII.

Notations: The following notation is used throughout this paper: $\mathbb{R}$ and $\mathbb{C}$ are the field of real and complex numbers. $\mathbf{A}$ represents a matrix, $\mathbf{a}$ and $\mathbf{a}^*$ are a vector and its conjugate, respectively. $\mathbf{a}_m$ is the mth column of $\mathbf{A}$ and $\mathbf{A}_{1:m}$ is a matrix containing the first m columns of $\mathbf{A}$. A_{mn} and $|A_{mn}|$ denote the (m, n) element of $\mathbf{A}$ and its magnitude. diag($\mathbf{A}_1$, $\mathbf{A}_2$, ..., $\mathbf{A}_K$)

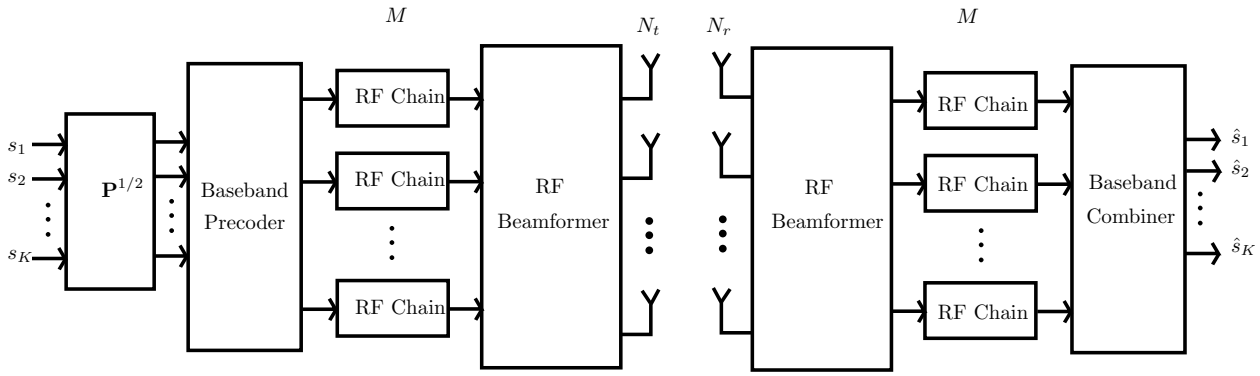

Fig. 1. The block diagram of a hybrid beamformer for a point-to-point scenario.

is a diagonal matrix with $\mathbf{A}_1$, $\mathbf{A}_2$, ..., $\mathbf{A}_K$ on its diagonal. $\mathbf{A}^{-1}$, det($\mathbf{A}$), $\|\mathbf{A}\|$, $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{\mathrm{H}}$, trace($\mathbf{A}$) denote inverse, determinant, Frobenius norm, transpose, Hermitian and trace of $\mathbf{A}$, respectively. $\mathcal{RN}(\mathbf{a}, \mathbf{A})$ and $\mathcal{CN}(\mathbf{a}, \mathbf{A})$ present a random vector of real and complex Gaussian distributed elements with expected value $\mathbf{a}$ and covariance matrix $\mathbf{A}$. Finally, $\mathbf{0}_{m\times 1}$, $\mathbf{1}_{m\times 1}$ and $\mathbf{I}_m$ are a vector of m zeros, m ones and an $m \times m$ identity matrix, respectively.

II. System Model

In a point-to-point MIMO communication system, the transmitter and the receiver are equipped with N_{t} and N_{r} antennas, respectively. The transmitter sends a vector $\mathbf{s} \in \mathbb{C}^{K\times 1}$ of K symbols to the receiver where $\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}] = \mathbf{I}_K$. The transmit signal vector becomes $\mathbf{x} = \sqrt{\frac{P_{\mathrm{t}}}{\Gamma_{\mathrm{t}}}}\mathbf{F}\mathbf{P}^{1/2}\mathbf{s}$, where P_{t} is the total transmit power, $\mathbf{P} \in \mathbb{R}^{K\times K}$ is a diagonal power allocation matrix with $\sum_{k=1}^{K} P_{kk} \leq 1$, $\mathbf{F} \in \mathbb{C}^{N_{\mathrm{t}}\times K}$ is the precoder matrix and $\Gamma_{\mathrm{t}} = \mathrm{trace}(\mathbf{F}^{\mathrm{H}}\mathbf{F})/K$ is a normalization factor such that $\|1/\sqrt{\Gamma_{\mathrm{t}}}\mathbf{F}\|^2 = K$. Let $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{r}}\times N_{\mathrm{t}}}$ and $\mathbf{y} \in \mathbb{C}^{N_{\mathrm{r}}\times 1}$ denote the normalized channel matrix and the received signal vector. Assuming the noise vector at the receiver antennas $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_{N_{\mathrm{r}}\times 1}, \sigma_z^2\mathbf{I}_{N_{\mathrm{r}}})$ has independent and identically distributed (i.i.d.) elements with variance σ_z^2, the channel input-output relationship is expressed as $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$. Applying the combiner matrix $\mathbf{W} \in \mathbb{C}^{N_{\mathrm{r}}\times K}$ at the receiver, the input $\hat{\mathbf{s}} \in \mathbb{C}^{K\times 1}$ to the detector is

$$\hat{\mathbf{s}} = \sqrt{\frac{P_{\mathrm{t}}}{\Gamma_{\mathrm{t}}\Gamma_{\mathrm{r}}}}\mathbf{W}^{\mathrm{H}}\mathbf{H}\mathbf{F}\mathbf{P}^{1/2}\mathbf{s} + \sqrt{\frac{1}{\Gamma_{\mathrm{r}}}}\mathbf{W}^{\mathrm{H}}\mathbf{z}, \tag{1}$$

where $\Gamma_{\mathrm{r}} = \mathrm{trace}(\mathbf{W}^{\mathrm{H}}\mathbf{W})/K$ is a normalization factor such that $\|1/\sqrt{\Gamma_{\mathrm{r}}}\mathbf{W}\|^2 = K$.

The block diagram of a point-to-point communication system with a hybrid beamformer is shown in Fig. 1. A hybrid beamformer consists of a baseband precoder $\mathbf{F}_{\mathrm{B}} \in \mathbb{C}^{M\times K}$ connected through M RF chains to the RF beamformer $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{t}}\times M}$ such that $\mathbf{F} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{B}}$. The elements of the RF beamformer are either analog or digital B-bit resolution phase shifters as

$$\mathbf{F}_{\mathrm{RF}} = \begin{pmatrix} \mathrm{e}^{j\theta_{11}} & \mathrm{e}^{j\theta_{12}} & \dots & \mathrm{e}^{j\theta_{1M}} \\ \mathrm{e}^{j\theta_{21}} & \mathrm{e}^{j\theta_{22}} & \dots & \mathrm{e}^{j\theta_{2M}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{e}^{j\theta_{N_{\mathrm{t}}1}} & \mathrm{e}^{j\theta_{N_{\mathrm{t}}2}} & \dots & \mathrm{e}^{j\theta_{N_{\mathrm{t}}M}} \end{pmatrix}, \quad \forall\, \theta_{n_{\mathrm{t}}m} \in \Theta,\ n_{\mathrm{t}} \in \{1,\ ...,\ N_{\mathrm{t}}\},\ m \in \{1,\ ...,\ M\}, \tag{2}$$

where $\Theta = [0, 2\pi]$ for analog phase shifters and $\Theta = \{0,\ 2\pi/2^B,\ ...,\ (2^B-1)2\pi/2^B\}$ for digital phase shifters. For the sake of the notation simplicity, throughout the paper it is assumed that $n_{\mathrm{t}} \in \{1,\ ...,\ N_{\mathrm{t}}\}$, $m \in \{1,\ ...,\ M\}$ and $k \in \{1,\ ...,\ K\}$ and the number of the RF chains at the transmitter and receiver are equal to M. Similar notation is used for the hybrid beamformer at the receiver as $\mathbf{W} = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{B}}$ where $\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{r}}\times M}$ and $\mathbf{W}_{\mathrm{B}} \in \mathbb{C}^{M\times K}$. Finally, the system model (1) for the hybrid scenario becomes

$$\hat{\mathbf{s}} = \sqrt{\frac{P_{\mathrm{t}}}{\Gamma_{\mathrm{t}}\Gamma_{\mathrm{r}}}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{B}}\mathbf{P}^{1/2}\mathbf{s} + \sqrt{\frac{1}{\Gamma_{\mathrm{r}}}}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{z}. \tag{3}$$

In this paper, we derive an asymptotically optimal hybrid beamformer for a narrowband system under rich and sparse scattering channels under the assumption that $\mathrm{E}[\|\mathbf{H}\|^2] = N_{\mathrm{r}}N_{\mathrm{t}}$. Rayleigh fading with i.i.d. elements $H_{n_{\mathrm{r}}n_{\mathrm{t}}} \sim \mathcal{CN}(0,1)$ is employed to model the rich scattering channel.

A geometry based model with $L \ll \min(N_{\mathrm{t}}, N_{\mathrm{r}})$ multipath components is applied for the sparse scattering scenario. In this case, the channel matrix is expressed as [16]

$$\mathbf{H} = \sqrt{\frac{N_{\mathrm{t}}N_{\mathrm{r}}}{L}}\sum_{l=1}^{L}\beta_l\mathbf{a}_{\mathrm{r}}(\phi_{\mathrm{r}l})\mathbf{a}_{\mathrm{t}}^{\mathrm{H}}(\phi_{\mathrm{t}l}), \tag{4}$$

where $\beta_l \sim \mathcal{CN}(0,1)$ is the multipath coefficient, $\phi_{\mathrm{t}l}$ and $\phi_{\mathrm{r}l}$ are angle-of-departure and angle-of-arrival of the lth multipath. Without loss of generality, it is assumed that $|\beta_1| \geq |\beta_2| \geq ... \geq |\beta_L|$. The steering vector $\mathbf{a}_u(\phi_{ul})$, $\forall u \in \{\mathrm{t},\ \mathrm{r}\}$, for linear arrays is expressed as

$$\mathbf{a}_u(\phi_{ul}) = \frac{1}{\sqrt{N_u}}(1, \mathrm{e}^{\frac{j2\pi d_u}{\lambda}\cos(\phi_{ul})}\ ...,\ \mathrm{e}^{\frac{j2\pi d_u}{\lambda}(N_u-1)\cos(\phi_{ul})})^{\mathrm{T}} \tag{5}$$

where $\phi_{ul} \in [0,\ \pi]$, λ is the wavelength and d_u is the antenna spacing [16]. In the rest of this paper, it is assumed that the transmitter and the receiver are equipped with linear arrays with $d_u = \lambda/2$.

III. Problem Statement and Motivation

The optimal beamforming and power allocation for a fully digital point-to-point system is achieved by singular value decomposition (SVD) and waterfilling. The SVD factorizes the channel matrix as $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$ where the columns of $\mathbf{V} \in \mathbb{C}^{N_{\mathrm{t}} \times N_{\mathrm{t}}}$ and $\mathbf{U} \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{r}}}$ are the right and left singular vectors of $\mathbf{H}$, and the diagonal elements of $\mathbf{\Sigma} \in \mathbb{R}^{N_{\mathrm{r}} \times N_{\mathrm{t}}}$ are the singular values of $\mathbf{H}$. For a full-ranked $\mathbf{H}$, the capacity of the MIMO channel at high SNR grows linearly with $\min(N_{\mathrm{t}},\ N_{\mathrm{r}})$ when $K = \min(N_{\mathrm{t}},\ N_{\mathrm{r}})$ streams are transmitted over the channel [17]. When $K \leq \min(N_{\mathrm{t}},\ N_{\mathrm{r}})$, the maximum achievable rates are derived by setting the combining and precoding matrices based on thin-SVD as $\mathbf{W}_{\mathrm{d}} = \mathbf{U}_{1:K}$ and $\mathbf{F}_{\mathrm{d}} = \mathbf{V}_{1:K}$ [11]. In this case, $\Gamma_{\mathrm{t}} = \Gamma_{\mathrm{r}} = 1$ and the capacity of a point-to-point system with K streams over $\mathbf{H}$ with Gaussian entries s_k is [11]

$$\begin{aligned} C = \max I(\mathbf{s};\hat{\mathbf{s}}) &= \max_{\text{trace}(\mathbf{P})\leq 1} \log_2 \det\Big(\mathbf{I}_K + \frac{P_{\mathrm{t}}}{\sigma_z^2}\mathbf{R}_{\mathrm{n}}^{-1}\mathbf{W}_{\mathrm{d}}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{d}}\mathbf{P}\mathbf{F}_{\mathrm{d}}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{W}_{\mathrm{d}}\Big) \\ &= \max_{\sum_{k=1}^{K} P_{kk}\leq 1} \sum_{k=1}^{K} \log_2(1 + P_{\mathrm{t}}P_{kk}\sigma_{kk}^2/\sigma_z^2) \end{aligned} \tag{6}$$

where $I(\mathbf{s};\hat{\mathbf{s}})$ is the mutual information between $\mathbf{s}$ and $\hat{\mathbf{s}}$, $\mathbf{R}_{\mathrm{n}} = \frac{1}{\Gamma_{\mathrm{r}}}\mathbf{W}^{\mathrm{H}}\mathbf{W} = \frac{1}{\Gamma_{\mathrm{r}}}\mathbf{W}_{\mathrm{d}}^{\mathrm{H}}\mathbf{W}_{\mathrm{d}} = \mathbf{I}_K$, σ_{kk}^2 are the ordered eigenvalues of $\mathbf{H}\mathbf{H}^{\mathrm{H}}$ and the optimal P_{kk} is derived by waterfilling. In this case, the capacity growth at high SNR is proportional to K. It should be noted that if the channel is rank-deficient it is not possible to transmit more than rank($\mathbf{H}$) symbols. When the hybrid beamformers are used, the achievable rate is expressed as [17]

$$R = I(\mathbf{s};\hat{\mathbf{s}}) = \log_2 \det\Big(\mathbf{I}_K + \frac{\rho}{\Gamma_{\mathrm{t}}\Gamma_{\mathrm{r}}}\mathbf{R}_{\mathrm{n}}^{-1}\mathbf{W}_{\mathrm{B}}^{\mathrm{H}}\mathbf{W}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{B}}\mathbf{P}\mathbf{F}_{\mathrm{B}}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{B}}\Big), \tag{7}$$

where $\rho = \frac{P_{\mathrm{t}}}{\sigma_z^2}$ is a measure of link signal-to-noise ratio (SNR).

One of the main challenges of designing hybrid beamformers is the joint design of the RF beamformers and baseband precoders/combiners considering the constant modulus constraint on the phase shifters. Designing $\mathbf{F}_{\mathrm{B}}$, $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{B}}$ and $\mathbf{W}_{\mathrm{RF}}$ to maximize (7) is a nonconvex problem and in general it is difficult to solve [8]. Due to the similarity between the hybrid beamformer matrices at the transmitter and the receiver, same design algorithms are applicable to both sides. Hence, the discussions and derivations during this paper are mostly focused on the hybrid beamformer at the transmitter. In this case, it is desired to maximize the mutual information $I(\mathbf{s};\mathbf{y})$ as

$$(\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}}, \mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}) = \arg\max_{\mathbf{F}_{\mathrm{B}},\mathbf{F}_{\mathrm{RF}}} I(\mathbf{s};\mathbf{y}) = \arg\max_{\mathbf{F}_{\mathrm{B}},\mathbf{F}_{\mathrm{RF}}} \log_2 \det\Big(\mathbf{I}_{N_{\mathrm{r}}} + \frac{\rho}{\Gamma_{\mathrm{t}}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{B}}\mathbf{P}\mathbf{F}_{\mathrm{B}}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\Big), \tag{8}$$

subject to (s.t.) $|F_{\mathrm{RF},n_t m}| = 1$, where $\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}}$ and $\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}$ are the optimal baseband precoding and RF beamforming matrices. It has been shown that based on some approximations, the optimization in (8) can be reformulated as [8]

$$(\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}}, \mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}) = \underset{\mathbf{F}_{\mathrm{B}}, \mathbf{F}_{\mathrm{RF}}}{\arg\min} \|\mathbf{F}_{\mathrm{d}} - \sqrt{\frac{1}{\Gamma_{\mathrm{t}}}} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{B}}\|, \quad \text{s.t. } |F_{\mathrm{RF},n_t m}| = 1. \tag{9}$$

A suboptimal joint baseband and RF design based on matching pursuit was proposed to solve (9) for a sparse scattering channel [8]. For a more general channel, including Rayleigh fading, this optimization problem can be approximated as a convex problem and a joint iterative suboptimal solution was proposed in [10]. Another approach to design the hybrid beamformer is to calculate $\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}$ at the first step, and then derive the $\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}}$ for the effective channel $\mathbf{H}_{\mathrm{e}} = \mathbf{H}\mathbf{F}_{\mathrm{RF}}$. Letting $\mathbf{x}_{\mathrm{B}} = \mathbf{F}_{\mathrm{B}}\mathbf{P}^{1/2}\mathbf{s}$, data-processing inequality indicates that [18]

$$I(\mathbf{s}; \mathbf{y}) \overset{(a)}{\leq} I(\mathbf{x}_{\mathrm{B}}; \mathbf{y}) \leq C. \tag{10}$$

where inequality (a) turns into equality when $\mathbf{F}_{\mathrm{B}} = \mathbf{V}_{\mathrm{e}}$ as $\mathbf{H}_{\mathrm{e}} = \mathbf{H}\mathbf{F}_{\mathrm{RF}} = \mathbf{U}_{\mathrm{e}}\mathbf{\Sigma}_{\mathrm{e}}\mathbf{V}_{\mathrm{e}}^{\mathrm{H}}$, and $\mathbf{P}$ is derived by waterfilling. It could be concluded that max $I(\mathbf{s}; \mathbf{y})$ only depends on the design of $\mathbf{F}_{\mathrm{RF}}$. In this case,

$$\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}} = \underset{\mathbf{F}_{\mathrm{RF}}}{\arg\max}\, I(\mathbf{s}; \mathbf{y}) = \underset{\mathbf{F}_{\mathrm{RF}}}{\arg\max}\, \log_2 \det\Big(\mathbf{I}_{N_{\mathrm{r}}} + \frac{\rho}{\Gamma_{\mathrm{t}}} \mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{V}_{\mathrm{e}}\mathbf{P}\mathbf{V}_{\mathrm{e}}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\Big), \quad \text{s.t. } |F_{\mathrm{RF},n_t m}| = 1, \tag{11}$$

where $\Gamma_{\mathrm{t}} = \mathrm{trace}(\mathbf{F}_{\mathrm{RF}}\mathbf{V}_{\mathrm{e}}\mathbf{V}_{\mathrm{e}}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}})/K = \mathrm{trace}(\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}})/K = N_{\mathrm{t}}$. The two-stage design of $\mathbf{F}_{\mathrm{B}}$ and $\mathbf{F}_{\mathrm{RF}}$ has been previously studied in [13]-[15]. However, the spectral efficiency based on these works depends on numerical calculations and it is not possible to derive the closed-form expression of the performance. Based on the two-stage approach, a virtually optimal hybrid beamforming and the closed-form expression of the spectral efficiency for a point-to-point system with large number of antennas under two specific channel scenarios are presented in the next section.

IV. Hybrid Beamforming for the Point-to-Point Scenario

In this section, an asymptotically optimal hybrid beamformer that maximizes the achievable rate in (8) is presented. Initially, based on some basic properties of the elements of the singular vectors, it will be shown that analog phase shifters with $K = M/2$ can achieve the performance of digital beamformers. It is notable that the analysis presented for this scenario is a modification of

the approach in [11]. Under this assumption the system is underperforming as the multiplexing gain is limited to $M/2$. In order to develop a hybrid beamforming algorithm that efficiently employs all the RF chains to transmit $K = M$ streams, some of the properties the singular vectors of $\mathbf{H}$ are investigated. Then, the hybrid beamforming solution for large antenna array systems with analog phase shifters are presented. For the case of $K < M < 2K$, a combination of the methods for $M = K$ and $M = 2K$, and its performance is discussed. When digital phase shifters are employed, a simple heuristic suboptimal solution and its performance lower-bound is presented. Finally, a discussion on the proposed method and a comparison with the state-of-the-art are provided at the end of the section.

Since $\mathbf{V}$ is a unitary matrix, $\mathbf{v}_k^{\mathrm{H}}\mathbf{v}_k = \sum_{n_{\mathrm{t}}=1}^{N_{\mathrm{t}}} |V_{n_{\mathrm{t}}k}|^2 = 1$ and $|V_{n_{\mathrm{t}}k}| \leq 1$. Thus $|V_{n_{\mathrm{t}}k}|$ is in the domain of the inverse cosine function. Hence,

$$\begin{aligned} |V_{n_{\mathrm{t}}k}|\mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}} &= \mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}} \cos\Big(\cos^{-1}(|V_{n_{\mathrm{t}}k}|)\Big) = \frac{\mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}}}{2}\mathrm{e}^{j\cos^{-1}(|V_{n_{\mathrm{t}}k}|)} + \frac{\mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}}}{2}\mathrm{e}^{-j\cos^{-1}(|V_{n_{\mathrm{t}}k}|)} \\ &= \frac{1}{2}\mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}+j\cos^{-1}(|V_{n_{\mathrm{t}}k}|)} + \frac{1}{2}\mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}-j\cos^{-1}(|V_{n_{\mathrm{t}}k}|)}. \end{aligned} \quad (12)$$

This means that two phase shifters and an adder at the RF are sufficient to produce $V_{n_{\mathrm{t}}k}$ when $M = K$ RF chains and $2MN_{\mathrm{t}}$ phase shifters are available. Alternative approach to the adders is employing $M = 2K$ RF chains and $2MN_{\mathrm{t}}$ phase shifters. In this case, $1/\sqrt{\Gamma_{\mathrm{t}}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}} = \mathbf{F}_{\mathrm{d}}$ is achieved by setting

$$F_{\mathrm{RF},n_{\mathrm{t}}k'}^{\mathrm{opt}} = \begin{cases} \mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}+j\cos^{-1}(|V_{n_{\mathrm{t}}k}|)} \text{ for } k' = 2k-1 \\ \mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}-j\cos^{-1}(|V_{n_{\mathrm{t}}k}|)} \text{ for } k' = 2k, \end{cases} \quad (13)$$

and $\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}} = \frac{1}{2}\mathrm{diag}(\mathbf{1}_{2\times 1},\ ...,\ \mathbf{1}_{2\times 1})$ and $\Gamma_{\mathrm{t}} = 1$. Hence, the maximum rate in (6) can be achieved with this design. In order to derive $\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}}$ and $\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}$ for $M = K$ scenario, further properties of the singular vectors are investigated in the following subsection.

A. Properties of the Channel Singular Vectors

The behaviors of the channel singular vectors for Rayleigh and geometry based models are presented in Theorem 1 and Lemma 1 in the following.

Theorem 1: If $H_{n_{\mathrm{r}}n_{\mathrm{t}}} \sim \mathcal{CN}(0,1)$ and $N_{\mathrm{t}} \to \infty$, $N_{\mathrm{r}} \to \infty$, then the elements of the singular vectors of $\mathbf{H}$ are i.i.d and follow $\sqrt{N_{\mathrm{t}}}V_{n_{\mathrm{t}}n_{\mathrm{t}}'}$, $\sqrt{N_{\mathrm{r}}}U_{n_{\mathrm{r}}n_{\mathrm{r}}'} \sim \mathcal{CN}(0,1)$, $\forall\ n_{\mathrm{t}}, n_{\mathrm{t}}' \in \{1,\ ...,\ N_{\mathrm{t}}\}$ and $\forall\ n_{\mathrm{r}}, n_{\mathrm{r}}' \in \{1,\ ...,\ N_{\mathrm{r}}\}$.

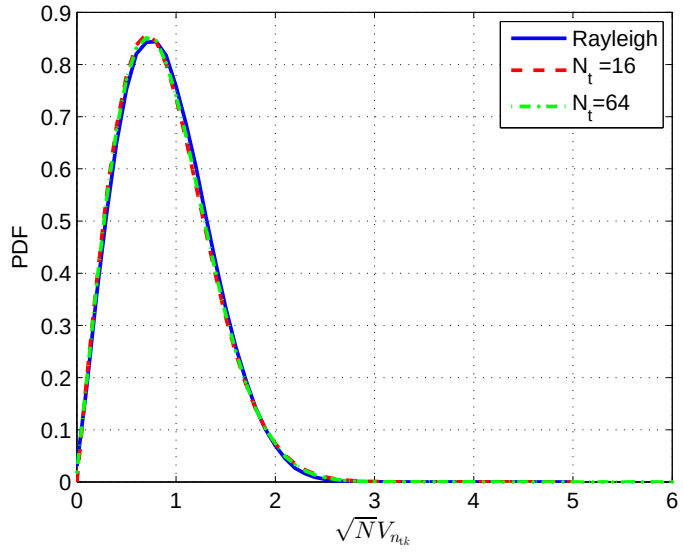

Fig. 2. Comparison between the probability density function (PDF) of $\sqrt{N_{\rm t}}|V_{n_{\rm t}k}|$ when it follows Rayleigh distribution with parameter $\sigma_{\rm R} = \frac{1}{\sqrt{2}}$, simulation results for the PDF of $\sqrt{N_{\rm t}}|V_{n_{\rm t}k}|$ for a Rayleigh fading channel over 1000 realizations with $N_{\rm t} = N_{\rm r} = 16$ and $N_{\rm t} = N_{\rm r} = 64$.

Proof: It is known that the left and right singular vectors of $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\rm H}$ are uniformly distributed on a complex $N_{\rm t}$-hypersphere and a $N_{\rm r}$-hypersphere with radius 1 [19]. As a result, $\sqrt{N_{\rm t}}\mathbf{v}_{n_{\rm t}}$ and $\sqrt{N_{\rm r}}\mathbf{u}_{n_{\rm r}}$ are uniformly distributed on the surface of a $N_{\rm t}$ and $N_{\rm r}$ dimensional hyperspheres with radius $\sqrt{N_{\rm t}}$ and $\sqrt{N_{\rm r}}$. Moreover, the coordinates of a randomly chosen point according to a uniform distribution on an N-hypersphere of radius $\sqrt{N}$ are i.i.d. with $\mathcal{CN}(0,1)$ when $N \to \infty$ [20]. Hence, the elements of $\sqrt{N_{\rm t}}\mathbf{v}_{n_{\rm t}}$ and $\sqrt{N_{\rm r}}\mathbf{u}_{n_{\rm r}}$ are i.i.d. with $\mathcal{CN}(0,1)$. $\square$

Remark: As far as the authors are aware, the distribution of the elements of the singular vectors of matrix $\mathbf{H}$, when $H_{n_{\rm r}n_{\rm t}} \sim \mathcal{CN}(0,1)$ for $N_{\rm t} \to \infty$ and $N_{\rm r} \to \infty$, has not been previously reported in the literature, although the pieces of the proof have been available for a long time and they have been studied by different researcher such as Love and Spruill [19], [20].

The real and imaginary parts of random variables with $\mathcal{CN}(0,1)$ are distributed as $\mathcal{RN}(0,\frac{1}{2})$ [16]. Hence, $\sqrt{N_{\rm t}}|V_{n_{\rm t}k}|$ has a Rayleigh distribution with parameter $\sigma_{\rm R} = \frac{1}{\sqrt{2}}$ and its expected value is $\sigma_{\rm R}\sqrt{\frac{\pi}{2}}$ [21]. Fig. 2 shows that the Rayleigh distribution can provide a good approximation even for a finite $N_{\rm t} \in \{16,\ 64\}$. The properties of the sparse scattering channels are described in the following lemma.

Lemma 1 [22]: For a geometry based channel model with $N_{\mathrm{t}} \to \infty$ and $N_{\mathrm{r}} \to \infty$, the relationship between the singular and steering vectors is expressed as $\mathbf{v}_l = \mathbf{a}_{\mathrm{t}}(\phi_{tl})$ and $\mathbf{u}_l = \mathbf{a}_{\mathrm{r}}(\phi_{\mathrm{r}l})$, $\forall\, l \in \{1,\, ...,\, L\}$.

B. Hybrid Beamforming for $M = K$ Scenario

The proposed hybrid beamformer when $M = K$ is presented in the following lemma.

Lemma 2: The asymptotically optimal solution $\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}$ and $\mathbf{W}_{\mathrm{RF}}^{\mathrm{opt}}$ to the optimizations in (7), (8) and (11) for large N_{t} and N_{r} with $M = K$ and analog phase shifters is $F_{\mathrm{RF},n_{\mathrm{t}}k}^{\mathrm{opt}} = \mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}}$, $W_{\mathrm{RF},n_{\mathrm{r}}k}^{\mathrm{opt}} = \mathrm{e}^{j\angle U_{n_{\mathrm{r}}k}}$. In this case, the baseband precoder and combiner matrices become $\mathbf{F}_{\mathrm{B}} = \mathbf{W}_{\mathrm{B}} = \mathbf{I}_K$.

Proof: Refer to Appendix A. □

It was previously shown that for the geometry based channel models, (8) could be approximated by (9) which is equivalent to minimizing the Euclidean distance between $\mathbf{F}_{\mathrm{d}}$ and $1/\sqrt{N_{\mathrm{t}}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{B}}$ [8]. In Appendix B, it is proved that the proposed RF beamformer of Lemma 2 can be alternatively derived by

$$\underset{\mathbf{F}_{\mathrm{RF}}}{\text{minimize}} \quad \|\sqrt{\frac{1}{N_{\mathrm{t}}}}\mathbf{F}_{\mathrm{RF}} - \mathbf{F}_{\mathrm{d}}\|^2, \quad \text{s.t. } |F_{\mathrm{RF},n_{\mathrm{t}}k}| = 1. \tag{14}$$

In order to implement the hybrid beamformer of Lemma 2, the first K singular vectors and values of $\mathbf{H}$ should be initially calculated. Then, each phase shifter at the transmitter and the receiver is directly set to the phase of the corresponding element in the right and left singular vectors, respectively. Considering the impact of the RF beamformers, the baseband precoder and combiner matrices are equal to an identity matrix. Finally, the optimal allocated power to each symbol is derived by waterfillining. The performance of the proposed hybrid beamformer compared to C in (6) for $M = K$ and Rayleigh fading channel is expressed in the following lemma.

Lemma 3: For large N_{t} and N_{r} and at high SNR regime, the difference between the maximum rate C form (6) and the rate R_{C} achieved by the beamforming scheme of Lemma 2 for a Rayleigh channel is expressed as

$$C - R_C = -2K \log_2(\frac{\pi}{4}). \tag{15}$$

Proof: Refer to Appendix C. □

Lemma 3 indicates that the spectral efficiency achieved by the digital beamformers is $0.7K$ bits/Hz/s more than the performance of the hybrid beamformers when the number of antennas are large and the channel is modeled by Rayleigh fading.

When $K < M < 2K$, two RF chains are used per symbol to transmit in the direction of the singular vectors corresponding to the first $M - K$ singular values of the channel. In this case, the hybrid beamformer of $M = 2K$ is used. For the remaining $2K - M$ symbols, the hybrid beamformer of Lemma 2 is used. That is $2(M - K)$ RF chains are used to transmit $M - K$ symbols and each of the remaining $2K - M$ symbols are transmitted on one of the remaining RF chains. For example, assuming that $K = 3$ and $M = 5$, the baseband precoder becomes $\mathbf{F}_{\text{B}}^{\text{opt}} = \text{diag}(\frac{1}{2}\mathbf{1}_{2\times 1},\ \frac{1}{2}\mathbf{1}_{2\times 1},\ 1)$. Then, (13) is used to design the the RF beamforming vectors $\mathbf{f}_{\text{RF},1},\ \mathbf{f}_{\text{RF},2}$ according to $\mathbf{v}_1$, and $\mathbf{f}_{\text{RF},3},\ \mathbf{f}_{\text{RF},4}$ based on $\mathbf{v}_2$. Finally, $\mathbf{f}_{\text{RF},5}$ is adjusted based on $\mathbf{v}_3$ and Lemma 2. Similar approach can be also applied at the receiver side. In a general scenario that $K \leq M \leq 2K$, by following the results of Appendix A and C, it can be easily verified that (15) becomes

$$C - R_C = -2(2K - M)\log_2(\frac{\pi}{4}). \tag{16}$$

For example, letting $K = 3$ and $M = 5$, then $C - R_C = -2\log_2(\frac{\pi}{4})$. It should be noted that adding an extra RF chain at each side can increase the spectral efficiency by $-\log_2(\frac{\pi}{4})$. However, this improvement will also increase the system cost, complexity and power consumption.

For a geometry based channel, the singular vectors and the steering vectors become equal and the proposed algorithm will be translated into steering the beams towards the channel multipath components as proposed in [22]. Following a similar approach as in Appendix C, it can be easily shown that $1/\sqrt{N_{\text{t}}}\mathbf{a}_{\text{t}}(\phi_{\text{t},k})^{\text{H}}\mathbf{f}_{\text{RF}} = 1$ and $C - R_{\text{C}} = 0$. Hence, extra RF chains $M - K > 0$ will not improve the performance in such channels.

C. Digital Phase Shifters

Another challenge for designing hybrid beamformers is the discrete resolution of the phase shifters. When B-bit resolution phase shifters are employed, the search space for the optimum set of phases becomes $2^{BMN_{\text{t}}}$ which can be very large for large N_{t}. As an example, when there are $N_{\text{t}} = 64$, $M = 4$ and 2-bit resolution phase shifters, there are 2^{512} possible phase combination which is computationally expensive to search in the real-time applications. One way out is the

use a predefined set of phases known as RF codebooks [8], [9]. The disadvantage of the RF codebooks is that they are usually designed for a fixed type of channel such as sparse channels. The alternative approach to design the RF beamformer with discrete resolution phase shifters is rounding the phases as

$$\theta^{\mathrm{d}}_{n_{\mathrm{t}}k} = \arg\min_{\theta_{n_{\mathrm{t}}k}} |\angle F_{\mathrm{d},n_{\mathrm{t}}k} - \theta_{n_{\mathrm{t}}k}|, \quad \text{s.t. } \theta_{n_{\mathrm{t}}k} \in \{0,\ ...,\ (2^B-1)2\pi/2^B\}, \tag{17}$$

where $\theta^{\mathrm{d}}_{n_{\mathrm{t}},k}$ is the phase of $F_{\mathrm{RF},n,k}$. The lower-bound on the rate loss with this design is provided in the following lemma.

Lemma 4: The gap between R_{C} and the achievable rate R_{D} by the hybrid beamformer based on (17) with B-bit resolution digital phase shifters is bounded as

$$R_{\mathrm{C}} - R_{\mathrm{D}} \leq -K \log_2 \left(\cos^4(\frac{2\pi}{2^{B+1}}) \right). \tag{18}$$

Proof: Refer to Appendix D. □

Lemma 4 indicates that hybrid beamformers with analog phase shifters can achieve maximum $0.45K$ bits/Hz/s higher spectral efficiency compared to the scenario that digital phase shifters with $B = 3$ are employed. As hybrid beamformers target the transmission of a small number of symbols, the gains achieved by using analog phase shifters are negligible at high SNR regime. In addition, the low cost and computational complexity of the proposed scheme in (17) makes it an effective approach for practical applications.

D. Discussion and Comparison with the State-of-the-Art

In this paper, the analytical discussions are focused on asymptotically large antenna arrays. This is in contrast to the works in [8]–[10], [10], [13], [15] where the analysis are presented for limited number of antennas. The advantages of considering asymptotically large arrays are two-fold. Firstly, it facilitates the analysis to derive the virtually optimal hybrid beamformer and the closed-forms for the achievable spectral efficiency. Secondly, as it will be shown in section VII, the simulation results indicate that the analysis for the asymptoticly large array scenario provides a reliable estimate of achievable performance for scenarios with limited number of antennas.

One of the common approaches in the literature is to decompose the unconstrained thin-SVD based beamformer matrix into RF beamformer and baseband precoder matrices, [8]–[10].

The computational complexity of the rank-M thin-SVD of $\mathbf{H}$ is $O(N_{\mathrm{t}}N_{\mathrm{r}}M)$ for $M \ll \sqrt{N_{\mathrm{t}}N_{\mathrm{r}}}$ [23]. The state-of-the-art hybrid beamformers that require a second round of computations to decompose $\mathbf{F}_{\mathrm{d}}$ into $\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}}$ and $\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}$ can cause high computational delay and complexity [8], [9]. An iterative algorithm can be used to solve the optimization problem in (8), however, the iterative algorithm renders a high computational cost and delay [10], [13], [15]. For example, the complexity of the hybrid beamformer in [14] is $O(\max(N_{\mathrm{t}}, N_{\mathrm{r}})^2 \min(N_{\mathrm{t}}, N_{\mathrm{r}}))$.

Compared to the state-of-the-art, the proposed hybrid beamformer of lemma 2 is faster and it is virtually the optimal scheme for the systems with large N_{t}, N_{r} operating in Rayleigh and sparse channels. The computational complexity of the proposed scheme is equal to the complexity of rank-M thin-SVD as $O(N_{\mathrm{t}}N_{\mathrm{r}}M)$. In addition, the closed-form expressions of the achievable rates are derived which to best of the authors' knowledge was not previously reported.

V. Multiuser Scenario

In the downlink scenario, the base station with N_{t} antennas transmits K symbols $\mathbf{s} \in \mathbb{C}^{K\times 1}$ to K single antenna mobile stations where $\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}] = 1/K\mathbf{I}_K$. In this scenario, it is assumed that the base station has perfect CSI and the users cannot collaborate. The total transmit power and the wireless channel matrix are denoted as P_{t} and $\mathbf{H} \in \mathbb{C}^{K\times N_{\mathrm{t}}}$, respectively. The transmit vector is expressed as $\mathbf{x} = \sqrt{P_{\mathrm{t}}/\Gamma_{\mathrm{t}}}\mathbf{F}\mathbf{s}$ where $\mathbf{F}$ is the precoding matrix and

$$\Gamma_{\mathrm{t}} = \mathrm{E}\Big[\mathrm{trace}(\mathbf{F}\mathbf{s}\mathbf{s}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}})\Big] = \mathrm{trace}\Big(\mathbf{F}\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}]\mathbf{F}^{\mathrm{H}}\Big) = \mathrm{trace}(\mathbf{F}\mathbf{F}^{\mathrm{H}})/K \tag{19}$$

is a power normalization factor. The channel output vector is $\mathbf{y} = (y_1, \ ..., \ y_K)^{\mathrm{T}}$ where y_k is the received signal at kth mobile station. The system input-output relation is expressed as $\mathbf{y} = \sqrt{\frac{P_{\mathrm{t}}}{\Gamma_{\mathrm{t}}}}\mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{z}$, where $\mathbf{z} = (z_1, \ ..., \ z_K)^{\mathrm{T}}$, $\mathrm{E}[\mathbf{z}\mathbf{z}^{\mathrm{H}}] = \sigma_z^2\mathbf{I}_K$ contains the receiver noise. The optimal sum-rate capacity of $\mathbf{H}$ is derived by [24]

$$C_{\mathrm{sum}}(P_{\mathrm{t}}, \mathbf{H}) = \max_{\mathrm{trace}(\mathbf{P})\leq 1} \log_2 \det\Big(\mathbf{I}_K + \frac{P_{\mathrm{t}}}{\sigma_z^2}\mathbf{P}\mathbf{H}\mathbf{H}^{\mathrm{H}}\Big), \tag{20}$$

where $\mathbf{P}$ is the power allocation matrix. In general, the capacity of the broadcast channels is derived by dirty paper coding which is difficult to implement [17]. Hence, in practice the suboptimal linear precoding algorithms with low complexity such as zero-forcing (ZF) are preferred. It has been shown that the performance of ZF converges to optimal sum-capacity for the Rayleigh channel when N_{t} goes large [1]. For the hybrid structure, the vector of the

received signals becomes $\mathbf{y} = \sqrt{\frac{P_t}{\Gamma_t}}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_B\mathbf{s}+\mathbf{z}$. In the following lemma, we present the virtually optimal hybrid beamformer and its performance, achievable sum-rate with respect to the sum-rate capacity of $\mathbf{H}$, for multiuser MIMO scenario when the channel is modeled by Rayleigh fading.

Lemma 5: The asymptotically optimal hybrid beamformer for the multiuser scenario with Rayleigh channel and in the high SNR regime consists of $\mathbf{F}_{RF}^{opt}$ from Lemma 2, and $\mathbf{F}_B^{opt} = (\mathbf{H}\mathbf{F}_{RF}^{opt})^{-1}$. In this case, the difference between the sum-capacity C_{sum} and the maximum achievable sum-rate R_{sum} at high SNR is

$$C_{sum}(P_t, \mathbf{H}) - R_{sum}(P_t, \frac{1}{\sqrt{\Gamma_t}}\mathbf{H}\mathbf{F}_{RF}^{opt}\mathbf{F}_B^{opt}) = -K\log_2(\pi/4). \tag{21}$$

Proof: Refer to Appendix E. □

The difference between right hand side of (15) and (21) is a scalar number 2. This factor comes from the fact that the transmitter and the receiver in the point-to-point system are equipped with hybrid beamformer, and the losses imposed by the RF beamformer should be counted at both sides.

For the case of the geometry based channels, the proposed RF beamformer in Lemma 2 is still asymptotically the optimal beamformer as it is shown in Appendix E. In order to achieve the maximum achievable rate, nonlinear precoding schemes should be used at the baseband precoder. The performance of ZF baseband precoder for sparse channels and the multiantenna multiuser scenario considering the impact of imperfect CSI on the system performance is investigated in [12]. Under the assumption of single antenna users, sparse channel, a base station with a linear array and ZF baseband precoder, the beamformer of Lemma 5 and the algorithm in [12] will result in the same performance. However, the hybrid beamformer in [12] is not applicable to Rayleigh channels due to employing a special RF codebook. The RF beamformer of Lemma 5, however, is applicable to both rich and sparse channels, and it is adaptable to different scenarios.

VI. Hybrid Beamforming with Phase Shifter Selection for Rayleigh Channel

In the previous sections, it was shown that the spectral efficiency achieved by the hybrid beamformers with MN_t phase shifters is comparable to the performance of the digital beamformers. However, the power consumption in the phase shifter network can be significant when large number of antennas are employed. Additionally, it is expected that the elements of $\mathbf{F}_{RF}$ corresponding to the elements of $\mathbf{F}_d$ with smaller amplitudes have less impact on the performance

of the beamformer. In this case, there are phase shifters with insignificant contribution to spectral efficiency although they consume the same amount of power. For this reason, a novel phase shifter selection scheme that turns off those shifters according to a predefined threshold is proposed in this section. The advantages of such an approach are twofold. Firstly it can improve the spectral efficiency as more power will be transmitted through the phase shifters with more contribution. Secondly, it can reduce the power consumption of the phase shifter network. In general, switches consume less power compared to phase shifters. The power consumption of each of the phase shifters P_{PS} and switches P_{S} at 2.4 GHz are reported as $28.8 \leq P_{\text{S}} \leq 152$ mW [25] and $0 < P_{\text{S}} \leq 15$ mW [26]. By switching off $\beta\%$ of the phase shifters, the total consumed power in the RF beamformer becomes $P_{\text{PSN}} = MN_{\text{t}}((1-\beta/100)P_{\text{PS}} + P_{\text{S}})$. For example, if $P_{\text{PS}} = 111$ mW, $P_{\text{S}} = 1$ mW, $M = 4$ and $N_{\text{t}} = 64$, the consumed power in the RF beamformer with all the phase shifters in operation (without switches) and the scenario that $\beta = 50$ are $MN_{\text{t}}P_{\text{PS}} = 28.4$ W and $P_{\text{PSN}} = 14.4$ W, respectively. That is close to 50% power saving in the phase shifter network. In this case, $\mathbf{F}_{\text{RF}}$ can be set as

$$F_{\text{RF},n_{\text{t}}k} = \begin{cases} 0, & \sqrt{N_{\text{t}}}|V_{n_{\text{t}}k}| \leq \alpha, \\ \text{e}^{j\angle V_{n_{\text{t}}k}}, & \alpha < \sqrt{N_{\text{t}}}|V_{n_{\text{t}}k}|, \end{cases} \tag{22}$$

where α is the threshold level. In Appendix F, it is shown that the relationship between α and β of is expressed as $\alpha = \sqrt{-\ln(1-\beta/100)}$. In addition, when the RF beamformer is set according to (22) the baseband precoder and combiner matrices become $\mathbf{F}_{\text{B}} = \mathbf{W}_{\text{B}} = \mathbf{I}_K$. In the following lemma, the closed-form expression for the performance of the proposed phase shifter selection algorithm is presented.

Lemma 6: In a Rayleigh channel and at high SNR and for large N_{t}, N_{r}, the spectral efficiency R_β achieved by the proposed phase shifter selection scheme, when $\beta\%$ of the phase shifters are switched off, compared to C is obtained from

$$C - R_\beta = 2K\log_2(1-\beta/100) - 4K\log\Big(\frac{\sqrt{\pi}}{2} + \alpha\text{e}^{-\alpha^2} - \frac{\sqrt{\pi}}{2}\text{erf}(\alpha)\Big), \tag{23}$$

where $\alpha = \sqrt{-\ln(1-\beta/100)}$.

Proof: Refer to Appendix F.

The closed-form (15) in Lemma 3 is a special case of (23) when $\beta = 0$. As a consequence of switching off the phase shifters that have smaller impact on the spectral efficiency, more power

can be allocated to the elements with higher impact. As a result, the choice of β can lead to a higher spectral efficiency as the power consumption in the RF beamformer is reduced. The relationship between β and R_β is studied in more detail by computer simulations in the next section.

VII. Simulation Results

In this section, the performance of the proposed hybrid beamforming schemes for the point-to-point and the multiuser scenarios operating in rich and sparse scattering channels is evaluated by Monte-Carlo simulations. The performance metric is average spectral efficiency over 1000 independent channel realizations and it is assumed that $M = K = 4$. In this paper, all the closed-form expressions were derived for the scenario that $N_{\rm t}, N_{\rm r} \to \infty$. To obtain the appropriate assumption on the number of the antenna elements for the simulations, the behavior of the hybrid beamformer with respect to $N_{\rm t}$ and $N_{\rm r}$ is first analyzed. In the following, the superscript "A"is used to distinguish the analytical results of the Lemmas 3, 4, 5 and 6 from the performance of the proposed schemes derived by the Monte-Carlo simulations. For example, the analytical spectral efficiency by the hybrid beamformer in Lemma 2 is expressed as $R_{\rm C}^{\rm A} = C - \Delta_{\rm C}$ where $\Delta_{\rm C} = \lim_{N_{\rm t},N_{\rm r}\to\infty} C - R_{\rm C}$. The performance of the hybrid beamformer of Lemma 2 with the closed-form expression of Lemma 3 are investigated with respect to the number of antennas, where it is assumed that $N_{\rm t} = N_{\rm r}$ and then $N_{\rm r} = 8$ for $N_{\rm t} \in \{8,\ 16,\ 32,\ 64,\ 128,\ 256,\ 512\}$. Figure 3 shows this performance $R_{\rm C}^{\rm A}$ compared to the simulation result, $R_{\rm C}$, for the Rayleigh fading channel whereas Fig. 4 presents these for the geometry based model. It is observed that $R_{\rm C}^{\rm A}$ and $R_{\rm C}$ converge for both channels when $N_{\rm t} = N_{\rm r}$ is large. For the Rayleigh fading channel, $R_{\rm C}^{\rm A}$ predicts slightly lower spectral efficiency compared to the results from simulations $R_{\rm C}$ when $N_{\rm r} = 8$ as shown in Fig. 3. On the other hand, it is observed from Fig. 4 that $R_{\rm C}^{\rm A}$ is always larger than $R_{\rm C}$ for the geometry based model as for this channel $C = R_{\rm C}^{\rm A}$.

Figure 5 shows the performance of the hybrid beamformer with digital phase shifters, denoted as $R_{\rm D}$, for a point-to-point system operating in rich scattering channel. It is observed that $R_{\rm C} - R_{\rm D}$ for $B = 2$ and $B = 3$ is 3.5 and 0.7 bits/s/Hz which is negligible compared to the high spectral efficiency achieved by large antenna arrays at high SNR. Hence, a simple rounding technique to set the discrete phases of the phase shifters with $B \geq 3$ can significantly simplify the calculations, and achieve a similar performance as analog phase shifters are employed. In

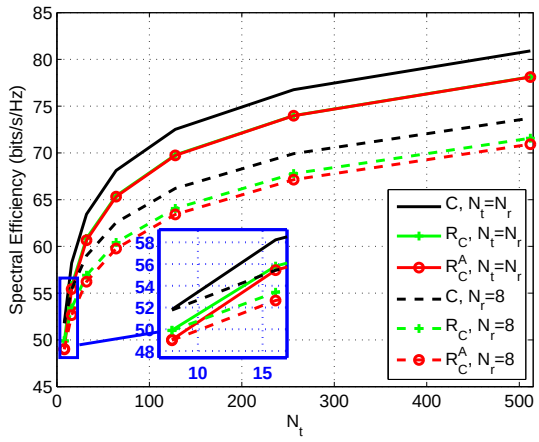

Fig. 3. C, $R_{\mathrm{C}}, R_{\mathrm{C}}^{\mathrm{A}}$ when the number of the antennas varies, $\rho = 34$ dB and Rayleigh channel.

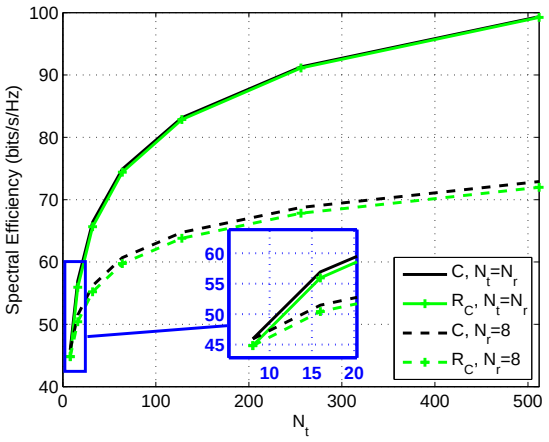

Fig. 4. C, $R_{\mathrm{C}}, R_{\mathrm{C}}^{\mathrm{A}}$ when the number of the antennas varies, $\rho = 34$ dB and geometry based channel with $L = 5$.

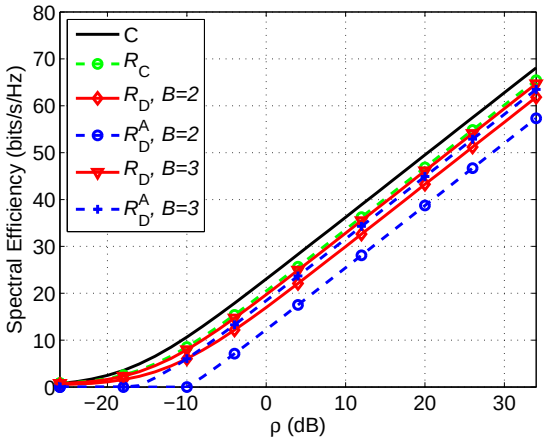

Fig. 5. Spectral efficiency achieved the hybrid beamformer with digital phase shifters based on (17) R_{D}, compared to the bound based on Lemma 4 $R_{\mathrm{D}}^{\mathrm{A}}$, R_{C} and C for Rayleigh channel.

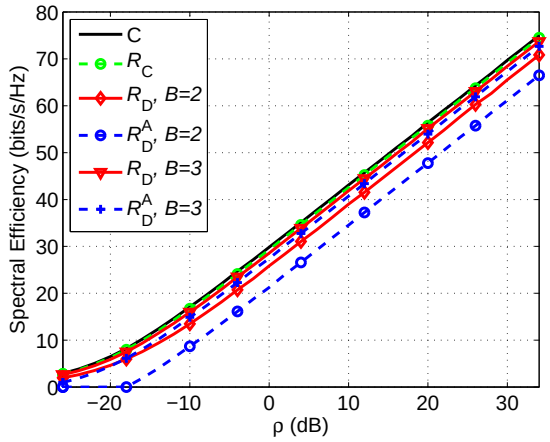

Fig. 6. Spectral efficiency achieved the hybrid beamformer with digital phase shifters based on (17) R_{D}, compared to the bound based on Lemma 4 $R_{\mathrm{D}}^{\mathrm{A}}$, R_{C} and C for geometry based channel with $L = 5$.

addition, the lower-bound of the spectral efficiency based on Lemma 4, denoted as $R_{\mathrm{D}}^{\mathrm{A}}$, provides a good approximation when $B \geq 3$. For example, when $B = 3$, $R_{\mathrm{D}} - R_{\mathrm{D}}^{\mathrm{A}}$ is 1.2 bits/s/Hz. Figure 6 presents a similar result for the sparse scattering channel. Figure 7 and Fig. 8 show the performance of the proposed algorithm in Lemma 2 compared to the state-of-the-art [8], [10], [14] for Rayleigh and geometry based channels. It is observed that the algorithm of [8] is not applicable to the Rayleigh fading channel, although it has a very good performance for the

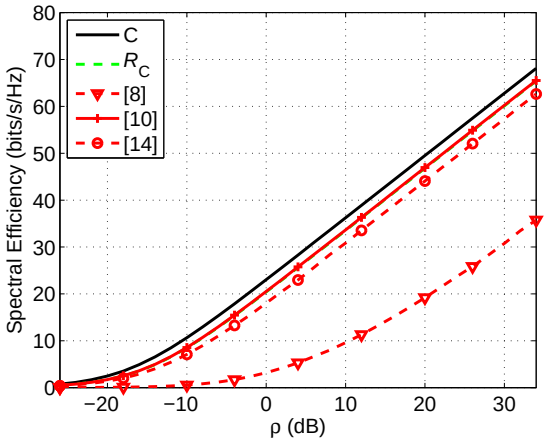

Fig. 7. Spectral efficiency achieved by the proposed algorithm compared to the state-of-the-art [8], [10], [14] when the wireless channel follows Rayleigh fading.

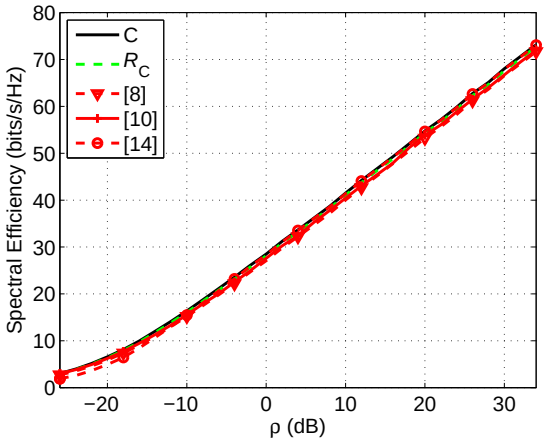

Fig. 8. Spectral efficiency achieved by the proposed algorithm compared to the state-of-the-art [8], [10], [14] when the wireless channel follows geometry based model with $L = 5$.

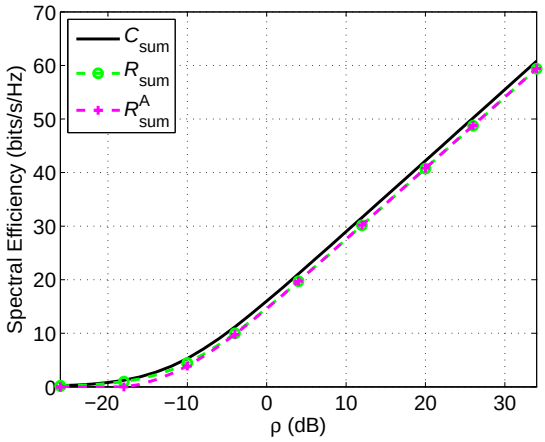

Fig. 9. Sum-rate achieved by ZF (digital beamforming) C_{sum}, the proposed hybrid beamformer for the multiuser scenario R_{sum} and the bound based on Lemma 5 $R^{\text{A}}_{\text{sum}}$ for Rayleigh fading channel.

sparse scattering channel. The performance of the iterative algorithms of [10] and [14] is similar to the proposed scheme for both channels.

For the downlink multiuser scenario with large number of antennas at the base station, ZF has been shown as the asymptotically optimal beamforming scheme in Rayleigh channels. Fig. 9 shows the achievable sum-rates by ZF with a digital beamformer and the proposed hybrid

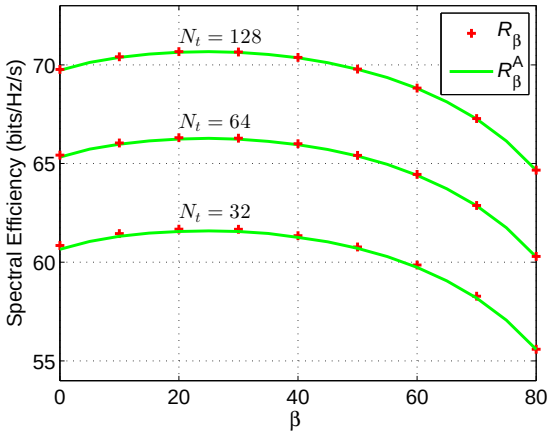

Fig. 10. Spectral efficiency achieved by the proposed phase shifter selection R_β, and the bound based on Lemma 6 R^{A}_β, $N_{\mathrm{t}} = N_{\mathrm{r}}$, $\rho = 34$ dB.

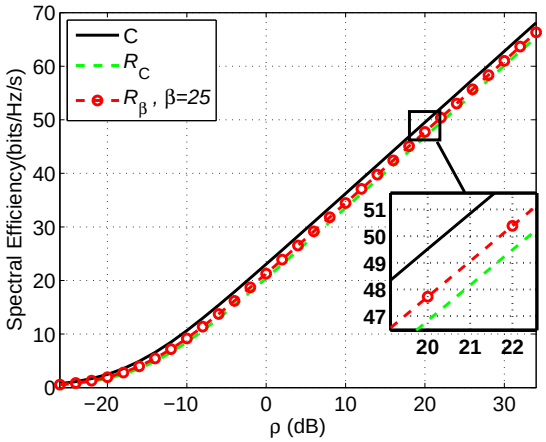

Fig. 11. Comparison of C, R_{C} and the spectral efficiency achieved by the proposed phase shifter selection with $\beta = 25$.

beamformer, denoted as C_{sum} as R_{sum}, when $N_{\mathrm{t}} = 64$ and $K = 4$. It is observed that the digital beamformer achieves 1.4 bit/s/Hz higher spectral efficiency than the hybrid beamformer as in Lemma 5.

Figure 10 shows the spectral efficiency achieved by the phase shifter selection scheme R_β compared to the closed-form based on Lemma 6, denoted as R^{A}_β, for Rayleigh fading channel and different values of β and N_{t}. It is observed that there is a good match between (23) and simulations. Compared to the scenario that all the phase shifters are in operation, the spectral efficiency can be improved when the phase shifter selection is applied with $0 < \beta < 50$. In addition, the maximum performance is achieved when β is around 25%. Finally, Fig. 11 presents the performance of the phase shifter selection scheme for $\beta = 25$ compared to R_{C} and C. It is observed that $C - R_{\beta=25}$ is around 33% smaller than $C - R_{\mathrm{C}}$. In addition, the spectral efficiency when all the phase shifters are in operation is almost equal to the case that $\beta\% = 50\%$ of them are turned off which results in 50% reduction in power consumption.

VIII. Conclusion and Future Works

In this paper, we derived the asymptotically optimal hybrid beamforming schemes to maximize the spectral efficiency for the point-to-point and multiuser systems with large antenna arrays, operating in rich and sparse scattering channels. The optimality of the solution was proved based

on the properties of the singular vectors of the channel matrix. The elements of these vectors have a complex Gaussian distribution for Rayleigh fading model, and the singular vectors are equal to the steering vectors of the channel matrix for the geometry based model. In addition, we derived the closed-form expressions for the spectral efficiency when the proposed hybrid beamformer is used. It was shown that the performance of the hybrid beamformer, employing phase shifters with resolution more than 2-bits, can approach the performance of a similar system with analog phase shifters. In order to reduce the power consumption in the RF beamformer, a novel phase shifter selection scheme was proposed. This approach can increase the spectral efficiency and reduce the power consumption when channel follows Rayleigh fading model. Simulation results indicate that spectral efficiency improves when up to 50% of the phase shifters are turned off.

The hybrid beamformer investigated in this work was developed and evaluated under certain assumptions such as perfect CSI, narrowband systems, no RF impairments, ideal phase shifters and switches. However, in order to integrate hybrid beamformers into practical systems, the impact of these parameters should be investigated.

APPENDIX A
PROOF OF LEMMA 2

Defining the positive semidefinite matrices $\mathbf{Q},\ \tilde{\mathbf{Q}} \in \mathbb{C}^{N_\mathrm{t}\times N_\mathrm{t}}$ as $\mathbf{Q} = \frac{1}{N_\mathrm{t}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{B}}\mathbf{P}\mathbf{F}_{\mathrm{B}}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}}$ and $\tilde{\mathbf{Q}} = \mathbf{V}^{\mathrm{H}}\mathbf{Q}\mathbf{V}$, the mutual information $I(\mathbf{s},\mathbf{y})$ is expressed as

$$\begin{aligned} I(\mathbf{s},\mathbf{y}) &= \log_2\det\Big(\mathbf{I}_{N_\mathrm{r}} + \rho\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{H}}\Big) = \log_2\det\Big(\mathbf{I}_{N_\mathrm{r}} + \rho\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{Q}\mathbf{V}\ \mathbf{\Sigma}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}\Big) \\ &= \log_2\det\Big(\mathbf{I}_{N_\mathrm{r}} + \rho\mathbf{\Sigma}\tilde{\mathbf{Q}}\mathbf{\Sigma}^{\mathrm{H}}\Big) = \log_2\det\Big(\mathbf{I}_{N_\mathrm{t}} + \rho\mathbf{\Sigma}^{\mathrm{H}}\mathbf{\Sigma}\tilde{\mathbf{Q}}\Big) \\ &\overset{(b)}{\leq} \log_2\Big(\prod_{n_\mathrm{t}=1}^{N_\mathrm{t}}(1+\rho\bar{\sigma}^2_{n_\mathrm{t}n_\mathrm{t}}\tilde{Q}_{n_\mathrm{t}n_\mathrm{t}})\Big), \end{aligned} \tag{24}$$

where $\bar{\sigma}^2_{n_\mathrm{t}n_\mathrm{t}}$ are the diagonal elements of $\mathbf{\Sigma}^{\mathrm{H}}\mathbf{\Sigma}$, and the inequality (b) comes from linear algebra as for any positive semidefinite matrix $\mathbf{A} \in \mathbb{C}^{N_\mathrm{t}\times N_\mathrm{t}}$, $\det(\mathbf{A}) \leq \prod_{n_\mathrm{t}} A_{n_\mathrm{t}n_\mathrm{t}}$. If $\tilde{\mathbf{Q}}$ is a diagonal matrix, then (b) in (24) turns into equality. Hence, the objective is to design $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{B}}$ such that they can diagonalize $\tilde{\mathbf{Q}}$. In order to analyze $\tilde{\mathbf{Q}}$, we investigate the behavior of the elements of $\mathbf{G} \in \mathbb{C}^{N_\mathrm{t}\times K}$, defiend as $\mathbf{G} = 1/\sqrt{N_\mathrm{t}}\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}$, when $\frac{1}{\sqrt{N_\mathrm{t}}}\mathbf{v}_{n_\mathrm{t}}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k} = 0$ and $\frac{1}{\sqrt{N_\mathrm{t}}}\mathbf{v}_{n_\mathrm{t}}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k} \neq 0$, $\forall\, n_\mathrm{t} \neq k$. In the first case that $\frac{1}{\sqrt{N_\mathrm{t}}}\mathbf{v}_{n_\mathrm{t}}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k} = 0\ \forall\, n_\mathrm{t} \neq k'$, it could be easily shown that all of the elements

of $\mathbf{G}$ except the G_{kk} become zero. Then, the last term in (24) can be written as

$$\log_2\Big(\prod_{n_{\mathrm{t}}=1}^{N_{\mathrm{t}}}(1+\rho\sigma^2_{n_{\mathrm{t}}n_{\mathrm{t}}}\tilde{Q}_{n_{\mathrm{t}}n_{\mathrm{t}}})\Big)=\log_2(\prod_{k=1}^{K}(1+\rho\sigma^2_{kk}\tilde{Q}_{kk})\Big). \tag{25}$$

On the other hand, if $\mathbf{F}_{\mathrm{B}}$ is a diagonal matrix, then $\mathbf{F}_{\mathrm{B}}\mathbf{P}\mathbf{F}_{\mathrm{B}}^{\mathrm{H}}$ will have the same property. As a result, $\tilde{\mathbf{Q}}=\mathbf{G}\mathbf{F}_{\mathrm{B}}\mathbf{P}\mathbf{F}_{\mathrm{B}}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}$ will also become a diagonal matrix since off-diagonal elements of $\mathbf{G}$ are zero. In addition, in (11), it was discussed that $\mathbf{F}_{\mathrm{B}}$ should be a unitary matrix to maximize the spectral efficiency. As $\mathbf{F}_{\mathrm{B}}$ is a diagonal and a unitary matrix, it could be concluded that $|F_{\mathrm{B},kk}|^2=1$. In this case, $\tilde{Q}_{kk}$ becomes

$$P_{kk}|G_{kk}|^2=\frac{P_{kk}}{N_{\mathrm{t}}}|\mathbf{v}_k^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}|^2=\frac{P_{kk}}{N_{\mathrm{t}}}|\sum_{n_{\mathrm{t}}=1}^{N_{\mathrm{t}}}V^*_{n_{\mathrm{t}}k}\mathrm{e}^{j\theta_{n_{\mathrm{t}}k}}|^2\overset{(c)}{\leq}\frac{P_{kk}}{N_{\mathrm{t}}}\Big|\sum_{n_{\mathrm{t}}=1}^{N_{\mathrm{t}}}|V_{n_{\mathrm{t}}k}|\Big|^2, \tag{26}$$

where the left hand side of (c) is maximized when all the elements of $\mathbf{v}_k$ are added constructively. In other words, (c) in (26) turns into equality if $\mathbf{F}_{\mathrm{RF}}=\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt1}}$ as

$$F^{\mathrm{opt1}}_{\mathrm{RF},n_{\mathrm{t}}k}=\mathrm{e}^{j\angle V_{n_{\mathrm{t}}k}}. \tag{27}$$

In the following, we analyze the impact of setting $\mathbf{F}_{\mathrm{RF}}=\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt1}}$ on the off-diagonal elements of $\mathbf{G}$ for rich and sparse scattering channels. For the Rayleigh channel, Theorem 1 expresses that the elements of singular vectors of the channel matrix are zero-mean i.i.d. random variables and their phases are uniformly distributed over $[0,\,2\pi]$. As a consequence of law of large numbers

$$\lim_{N_{\mathrm{t}}\to\infty}\frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{v}_{n_{\mathrm{t}}}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}=\lim_{N_{\mathrm{t}}\to\infty}\frac{1}{N_{\mathrm{t}}}\sum_{n'_{\mathrm{t}}=1}^{N_{\mathrm{t}}}\sqrt{N_{\mathrm{t}}}V^*_{n'_{\mathrm{t}}n_{\mathrm{t}}}\mathrm{e}^{j\angle V_{n'_{\mathrm{t}}k}}=\mathrm{E}[\sqrt{N_{\mathrm{t}}}V_{n_{\mathrm{t}}k}]=0, \tag{28}$$

for $n_{\mathrm{t}}\neq k$. For the geometry based model, Lemma 1 states that the RF precoder in (27) becomes $\mathbf{f}_{\mathrm{RF},k}=\sqrt{N_{\mathrm{t}}}\mathbf{v}_k$, hence $1/\sqrt{N_{\mathrm{t}}}\mathbf{v}_{n_{\mathrm{t}}}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}=\mathbf{v}_{n_{\mathrm{t}}}^{\mathrm{H}}\mathbf{v}_k=0,\ \forall\, n_{\mathrm{t}}\neq k$. As a result, it could be concluded that all of the elements of $\mathbf{G}$ except the diagonal elements become zero for both channels, when $\mathbf{F}_{\mathrm{RF}}=\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt1}}$. As a result, the choice of $\mathbf{F}_{\mathrm{RF}}=\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt1}}$ and a diagonal $\mathbf{F}_{\mathrm{B}}$, with $|F_{\mathrm{B},kk}|^2=1$, imposes (b) in (24) to turn into equality. Finally, $I(\mathbf{s},\mathbf{y})$ is maximized when the diagonal matrix $\mathbf{P}$ is calculated based on waterfilling.

It could be easily shown that when the hybrid beamformer at the receiver is also considered, by applying a similar RF beamformer at the receiver, $\sqrt{\frac{1}{N_{\mathrm{t}}N_{\mathrm{r}}}}\mathbf{W}_{\mathrm{RF}}^{\mathrm{opt}^{\mathrm{H}}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}$ becomes a diagonal matrix for both channels. In addition, $\mathbf{W}_{\mathrm{B}}$ will have a similar structure to $\mathbf{F}_{\mathrm{B}}$. Hence, $\mathbf{F}_{\mathrm{B}}=\mathbf{W}_{\mathrm{B}}=\mathbf{I}_K$, $\Gamma_{\mathrm{t}}=N_{\mathrm{t}}$, $\Gamma_{\mathrm{r}}=N_{\mathrm{r}}$ is the capacity achieving hybrid beamformer for both channels. $\square$

APPENDIX B
ALTERNATIVE DERIVATION OF THE RF BEAMFORMER IN LEMMA 2

Since $\mathbf{V}$ is a unitary matrix, $\|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{F}_{\mathrm{RF}}-\mathbf{F}_{\mathrm{d}}\|^2=\|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}-\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{d}}\|^2$. It could be easily verified that

$$\|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}-\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{d}}\|^2\geq\|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{F}_{\mathrm{d}}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}-\mathbf{I}_K\|^2,\quad\text{s.t. }|F_{\mathrm{RF},n_\mathrm{t}k}|^2=1. \tag{29}$$

The right hand side of the inequality can be reformulated as

$$\underset{\mathbf{F}_{\mathrm{RF}}}{\text{minimize}}\quad\sum_{k=1}^{K}\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}-1\Big|^2+\sum_{k'=1}^{K}\sum_{\substack{k=1\\k\neq k'}}^{K}\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k'}\Big|^2,\quad\text{s.t. }|F_{\mathrm{RF},n_\mathrm{t}k}|^2=1. \tag{30}$$

The cost function can be lower-bounded as

$$\begin{aligned}
&\min\Big(\sum_{k=1}^{K}\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}-1\Big|^2+\sum_{k'=1}^{K}\sum_{\substack{k=1\\k\neq k'}}^{K}\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k'}\Big|^2\Big) \qquad (31)\\
&=\min\Big(\sum_{k=1}^{K}\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}-1\Big|^2\Big)+\min\Big(\sum_{k'=1}^{K}\sum_{\substack{k=1\\k\neq k'}}^{K}\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k'}\Big|^2\Big)\\
&\overset{(d)}{\geq}\min\left(\sum_{k=1}^{K}\Big(\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}\Big|-1\Big)^2\right)+\min\Big(\sum_{k'=1}^{K}\sum_{\substack{k=1\\k\neq k'}}^{K}\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k'}\Big|^2\Big)\\
&\overset{(e)}{\geq}\sum_{k=1}^{K}\min\Big(\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}\Big|-1\Big)^2\overset{(f)}{=}\sum_{k=1}^{K}\left(\max\Big(\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}\Big|\Big)-1\right)^2,
\end{aligned}$$

where (f) comes from the fact that $\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}\Big|\leq 1$. Hence, the last term in (31) is minimized if

$$\underset{\mathbf{f}_{\mathrm{RF},k}}{\text{maximize}}\quad\Big|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}\Big|,\quad\text{s.t. }|F_{\mathrm{RF},n_\mathrm{t}k}|^2=1, \tag{32}$$

which is similar to (26) in Appendix (A). It was shown that $|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k'}|=0$, $\forall k\neq k'$ and $\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{f}_{\mathrm{d},k}^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}$ becomes a real and positive number when $F_{\mathrm{RF},n_\mathrm{t}k}^{\mathrm{opt}}=\mathrm{e}^{j\angle F_{\mathrm{d},n_\mathrm{t}k}}$. Hence, (e) and (d) turn into equality, and the cost function in (30) is minimized. Finally, (29) turns into equality and $\|\sqrt{\frac{1}{N_\mathrm{t}}}\mathbf{F}_{\mathrm{RF}}-\mathbf{F}_{\mathrm{d}}\|^2$ is minimized. □

APPENDIX C
PROOF OF OF LEMMA 3

As a result of Theorem 1, $\mathrm{E}[\sqrt{N_\mathrm{t}}|V_{n_\mathrm{t}k}|] = \mathrm{E}[\sqrt{N_\mathrm{r}}|U_{n_\mathrm{r}k}|] = \frac{\sqrt{\pi}}{2}$, and hence,

$$\lim_{N_\mathrm{t}\to\infty}\frac{1}{\sqrt{N_\mathrm{t}}}|\mathbf{v}_\mathrm{k}^\mathrm{H}\mathbf{f}_{\mathrm{RF},k}^{\mathrm{opt}}| = \lim_{N_\mathrm{r}\to\infty}\frac{1}{\sqrt{N_\mathrm{r}}}|\mathbf{u}_\mathrm{k}^\mathrm{H}\mathbf{w}_{\mathrm{RF},k}^{\mathrm{opt}}| = \lim_{N_\mathrm{t}\to\infty}\frac{1}{N_\mathrm{t}}\Big|\sum_{n_\mathrm{t}=1}^{N_\mathrm{t}}|\sqrt{N_\mathrm{t}}V_{n_\mathrm{t}k}|\Big| = \frac{\sqrt{\pi}}{2}. \tag{33}$$

Referring to the matrix $\mathbf{G} = 1/\sqrt{N_\mathrm{t}}\mathbf{V}^\mathrm{H}\mathbf{F}_\mathrm{RF}$ in Appendix A, $G_{kk} = \sqrt{\pi}/2$ and $G_{n_\mathrm{t}k} = 0$, $\forall n_\mathrm{t} \neq k$. Applying a similar RF beamformer at the receiver side, it could be easily verified that $\mathbf{R}_\mathrm{n} = 1/N_\mathrm{r}\mathbf{W}_\mathrm{RF}^\mathrm{H}\mathbf{W}_\mathrm{RF} = \mathbf{I}_K$. The spectral efficiency in (7) at high SNR becomes

$$\begin{aligned}
R_\mathrm{C} &= \lim_{N_\mathrm{t}\to\infty}\lim_{N_\mathrm{r}\to\infty}\log_2\det\Big(\frac{\rho}{N_\mathrm{t}N_\mathrm{r}}\mathbf{R}_\mathrm{n}^{-1}\mathbf{W}_\mathrm{B}^\mathrm{H}\mathbf{W}_\mathrm{RF}^\mathrm{H}\mathbf{H}\mathbf{F}_\mathrm{RF}\mathbf{F}_\mathrm{B}\mathbf{P}\mathbf{F}_\mathrm{B}^\mathrm{H}\mathbf{F}_\mathrm{RF}^\mathrm{H}\mathbf{H}^\mathrm{H}\mathbf{W}_\mathrm{RF}\mathbf{W}_\mathrm{B}\Big) \\
&= \lim_{N_\mathrm{t}\to\infty}\lim_{N_\mathrm{r}\to\infty}\log_2\det\Big(\frac{\rho}{N_\mathrm{t}N_\mathrm{r}}\mathbf{W}_\mathrm{RF}^\mathrm{H}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathrm{H}\mathbf{F}_\mathrm{RF}\mathbf{P}\mathbf{F}_\mathrm{RF}^\mathrm{H}\mathbf{V}\mathbf{\Sigma}\mathbf{U}^\mathrm{H}\mathbf{W}_\mathrm{RF}\Big) \\
&= \log_2\det\left(\Big(\frac{\pi}{4}\Big)^2\rho\mathbf{\Sigma'}^2\mathbf{P}\right) = \sum_{k=1}^{K}\log_2(\frac{\pi^2}{4^2}\rho P_{kk}\sigma_{kk}^2) \\
&= \sum_{k=1}^{K}\log_2(\rho P_{kk}\sigma_{kk}^2) + 2K\log_2(\frac{\pi}{4}),
\end{aligned} \tag{34}$$

where $\mathbf{\Sigma'} = \mathrm{diag}(\sigma_1^2,\ ...,\ \sigma_K^2)$. Considering that the first term in the last line is C in (6), the lemma is proved. □

APPENDIX D
PROOF OF OF LEMMA 4

In Appendix A, it was shown that the achievable rate depends on $\frac{1}{\sqrt{N_\mathrm{t}}}|\mathbf{v}_k^\mathrm{H}\mathbf{f}_{\mathrm{RF},k}|$. Letting $\delta_{n_\mathrm{t}k} = \theta_{n_\mathrm{t}k}^\mathrm{d} - \angle F_{\mathrm{d},n_\mathrm{t}k}$ where $\frac{-2\pi}{2^{B+1}} \leq \delta_{n_\mathrm{t}k} \leq \frac{2\pi}{2^{B+1}}$,

$$\begin{aligned}
\frac{1}{\sqrt{N_\mathrm{t}}}|\mathbf{v}_k^\mathrm{H}\mathbf{f}_{\mathrm{RF},k}| &= \frac{1}{\sqrt{N_\mathrm{t}}}\Big|\sum_{n_\mathrm{t}=1}^{N_\mathrm{t}}|V_{n_\mathrm{t}k}|\mathrm{e}^{-j\angle V_{n_\mathrm{t}k}}\mathrm{e}^{j\theta_{n_\mathrm{t}k}^*}\Big| = \frac{1}{\sqrt{N_\mathrm{t}}}\Big|\sum_{n_\mathrm{t}=1}^{N_\mathrm{t}}|V_{n_\mathrm{t}k}|\mathrm{e}^{j\delta_{n_\mathrm{t}k}}\Big| \\
&= \frac{1}{\sqrt{N_\mathrm{t}}}\left|\sum_{n_\mathrm{t}=1}^{N_\mathrm{t}}|V_{n_\mathrm{t}k}|\Big(\cos(\delta_{n_\mathrm{t}k}) + j\sin(\delta_{n_\mathrm{t}k})\Big)\right| \geq \frac{1}{\sqrt{N_\mathrm{t}}}\Big|\sum_{n_\mathrm{t}=1}^{N_\mathrm{t}}|V_{n_\mathrm{t}k}|\cos(\delta_{n_\mathrm{t}k})\Big| \\
&\geq \frac{1}{\sqrt{N_\mathrm{t}}}\cos(\frac{2\pi}{2^{B+1}})\sum_{n_\mathrm{t}=1}^{N_\mathrm{t}}|V_{n_\mathrm{t}k}|.
\end{aligned} \tag{35}$$

It could be easily shown that $1/\sqrt{N_\mathrm{t}}|\mathbf{v}_k^\mathrm{H}\mathbf{f}_{k'}| = 0 \forall k \neq k'$ holds for both channel models. Following a similar approach as in Appendix C, the rest of the proof is straight forward. □

APPENDIX E
PROOF OF LEMMA 5

Letting $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$ and $\mathbf{F}_{\mathrm{d}} = \mathbf{V}_{1:K}$, the sum-rate capacity of a multiuser broadcast channel can be expressed as [24]

$$C_{\mathrm{sum}}(P_{\mathrm{t}}, \mathbf{H}) = \max_{\mathrm{trace}(\mathbf{P})\leq 1} \log_2 \det\Big(\mathbf{I}_K + \frac{P_{\mathrm{t}}}{\sigma_z^2}\mathbf{P}\mathbf{H}\mathbf{H}^{\mathrm{H}}\Big) = \max_{\mathrm{trace}(\mathbf{P})\leq 1} \log_2 \det\Big(\mathbf{I}_K + \frac{P_{\mathrm{t}}}{\sigma_z^2}\mathbf{P}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}\Big) \tag{36}$$

$$= \max_{\mathrm{trace}(\mathbf{P})\leq 1} \log_2 \det\Big(\mathbf{I}_K + \frac{P_{\mathrm{t}}}{\sigma_z^2}\mathbf{P}\mathbf{U}\mathbf{\Sigma}_{1:K}^2\mathbf{U}^{\mathrm{H}}\Big) = C_{\mathrm{sum}}(P_{\mathrm{t}}, \mathbf{H}\mathbf{F}_{\mathrm{d}}),$$

where the last two equalities comes from the fact that $\mathbf{\Sigma}$ has only K nonzero elements and $\mathbf{\Sigma}_{1:K}^2 = \mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}} = \mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{d}}\mathbf{F}_{\mathrm{d}}^{\mathrm{H}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}}$. As the singular vectors of the channel are in the direction of the channel steering vectors, it can be easily concluded that the RF beamformer of Lemma 2 is virtually the optimal scheme for the sparse channel model.

For the Rayleigh channel employing $\mathbf{F}_{\mathrm{d}}$ as the RF beamformer is equivalent to relaxing the constant modulus constraint of the phase shifters. When $N_{\mathrm{t}} \to \infty$, the performance of ZF beamformer with $\mathbf{F}_{\mathrm{ZF}} = \mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}$ converges to the sum-capacity and the channel input-output relationship becomes

$$\mathbf{y} = \frac{1}{\sqrt{\Gamma_{\mathrm{t}}}}\mathbf{H}\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{s} + \mathbf{z} = \frac{1}{\sqrt{\Gamma_{\mathrm{t}}}}\mathbf{s} + \mathbf{z}. \tag{37}$$

In this case, Γ_{ZF} in (19) is

$$\Gamma_{\mathrm{t}} = \frac{1}{K}\mathrm{trace}(\mathbf{F}_{\mathrm{ZF}}\mathbf{F}_{\mathrm{ZF}}^{\mathrm{H}}) = \frac{1}{K}\mathrm{trace}\Big(\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\big((\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\big)^{\mathrm{H}}\mathbf{H}\Big) \tag{38}$$

$$= \frac{1}{K}\mathrm{trace}\Big(\mathbf{H}\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\big((\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\big)^{\mathrm{H}}\Big) = \frac{1}{K}\mathrm{trace}\Big((\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\Big).$$

The spectral efficiency achieved by ZF is expressed as

$$C_{\mathrm{sum}}(P_{\mathrm{t}}, \mathbf{H}) = C_{\mathrm{sum}}(P_{\mathrm{t}}, \sqrt{\Gamma_{\mathrm{t}}}\mathbf{H}\mathbf{F}_{\mathrm{ZF}}) = K\log_2(1+\rho) \tag{39}$$

$$= K\log_2(1 + \frac{P_{\mathrm{t}}\mathrm{E}[|s_k|^2]}{\Gamma_{\mathrm{t}}\sigma_z^2}) = K\log_2(1 + \frac{P_{\mathrm{t}}}{K\Gamma_{\mathrm{t}}\sigma_z^2}),$$

where ρ is the received SNR at the user side and $\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}] = 1/K\mathbf{I}_K$, respectively.

In addition, by applying ZF to the effective channel $\mathbf{H}_{\mathrm{e}} = \mathbf{H}\mathbf{F}_{\mathrm{d}} = \mathbf{U}\mathbf{\Sigma}_{1:K}$, the precoder matrix becomes $\mathbf{F}_{\mathrm{ZFe}} = \mathbf{H}_{\mathrm{e}}^{-1} = \mathbf{\Sigma}_{1:K}^{-1}\mathbf{U}^{\mathrm{H}}$. It should be noted that $\mathbf{F}_{\mathrm{d}}^{\mathrm{H}}\mathbf{F}_{\mathrm{d}} = \mathbf{I}_K$, and the rank of $\mathbf{H} \in \mathbb{C}^{K\times N_{\mathrm{t}}}$

is K and hence $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}}$ has only K nonzero elements. Then, the normalization factor Γ_{t} can be calculated as

$$\begin{aligned}\Gamma_{\mathrm{t}} &= \frac{1}{K}\mathrm{trace}(\mathbf{F}_{\mathrm{d}}\mathbf{F}_{\mathrm{ZFe}}\mathbf{F}_{\mathrm{ZFe}}^{\mathrm{H}}\mathbf{F}_{\mathrm{d}}^{\mathrm{H}}) = \frac{1}{K}\mathrm{trace}(\mathbf{F}_{\mathrm{ZFe}}\mathbf{F}_{\mathrm{ZFe}}^{\mathrm{H}}) = \frac{1}{K}\mathrm{trace}(\mathbf{\Sigma}_{1:K}^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{\Sigma}_{1:K}^{-1}) \\ &= \frac{1}{K}\mathrm{trace}(\mathbf{\Sigma}_{1:K}^{-2}) = \frac{1}{K}\mathrm{trace}\big((\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}})^{-1}\big) = \frac{1}{K}\mathrm{trace}\big((\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}})^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{U}\big) \\ &= \frac{1}{K}\mathrm{trace}\big((\frac{N_{\mathrm{t}}}{N_{\mathrm{t}}}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}})^{-1}\big) = \frac{1}{K}\mathrm{trace}\Big((\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\Big).\end{aligned} \tag{40}$$

As a consequence, $1/\sqrt{\Gamma_{\mathrm{t}}}\mathbf{H}\mathbf{F}_{\mathrm{ZF}} = 1/\sqrt{N_{\mathrm{t}}}\mathbf{H}\mathbf{F}_{\mathrm{d}}\mathbf{F}_{\mathrm{ZFe}}$ and

$$C_{\mathrm{sum}}(P_{\mathrm{t}}, \mathbf{H}) = C_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{\Gamma_{\mathrm{t}}}}\mathbf{H}_{\mathrm{e}}\mathbf{F}_{\mathrm{ZFe}}). \tag{41}$$

Hence, the asymptotically optimal hybrid beamforming scheme is derived when the constant modulus constraint at the RF beamformer is relaxed.

Since K is fixed and $N_{\mathrm{t}} \to \infty$, the array gain, and consequently, the received SNR also grows large. For MIMO channels with high SNR, in Theorem 3 of [27] and Theorem 2 of [28], it was shown that

$$\lim_{\rho\to\infty}\big[C(P_{\mathrm{t}}, \mathbf{H}) - C_{\mathrm{sum}}(P_{\mathrm{t}}, \mathbf{H})\big] = 0, \tag{42}$$

where $C(P_{\mathrm{t}}, \mathbf{H})$ is the capacity of the point-to-point system. Considering $C(P_{\mathrm{t}}, \mathbf{H}) = C(P_{\mathrm{t}}, \mathbf{U}^{\mathrm{H}}\mathbf{H})$, it could be concluded that

$$C_{\mathrm{sum}}(P_{\mathrm{t}}, \mathbf{U}^{\mathrm{H}}\mathbf{H}) = C(P_{\mathrm{t}}, \mathbf{U}^{\mathrm{H}}\mathbf{H}) = C(P_{\mathrm{t}}, \mathbf{H}) = C_{\mathrm{sum}}(P_{\mathrm{t}}, \mathbf{H}). \tag{43}$$

Let $R_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{H}\mathbf{F}_{\mathrm{RF}})$ denote the achievable sum-rate of multiuser scenario when the constant modulus is taken into account. Similar to (43), it could be easily verified that

$$\begin{aligned}R_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}) &= R_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}) \\ &= \max_{\mathrm{trace}(\mathbf{P})\le 1} \log_2 \det\Big(\mathbf{I}_K + \frac{P_{\mathrm{t}}}{N_{\mathrm{t}}\sigma_z^2}\mathbf{P}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}}\Big).\end{aligned} \tag{44}$$

Now, the RF beamformer that maximizes $R_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{H}\mathbf{F}_{\mathrm{RF}})$ is obtained by

$$\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}} = \arg\max_{\mathbf{F}_{\mathrm{RF}}} R_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}), \text{ s.t. } |F_{\mathrm{RF},n_{\mathrm{t}}k}| = 1. \tag{45}$$

Similar to Appendix A, $\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}$ of Lemma 2 that can diagonalize $\mathbf{P}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{H}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}}$ will also maximize $R_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{RF}})$ in (44). On the other hand, in Appendix C it is shown that

$\frac{1}{\sqrt{N_\mathrm{t}}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}} = \frac{\sqrt{\pi}}{2}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{d}} = \frac{\sqrt{\pi}}{2}\mathbf{H}_{\mathrm{e}}$. Additionally, in (41) it was discussed that $\mathbf{F}_{\mathrm{ZFe}}$ is asymptotically optimal for $\mathbf{H}_{\mathrm{e}}$, as a result

$$\begin{aligned} R_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}) &= \max_{\mathrm{trace}(\mathbf{P})\leq 1} \log_2 \det\Big(\mathbf{I}_K + \frac{P_{\mathrm{t}}}{N_{\mathrm{t}}\sigma_z^2}\mathbf{P}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}^{\mathrm{H}}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}}\Big) \\ &= \max_{\mathrm{trace}(\mathbf{P})\leq 1} \log_2 \det\Big(\mathbf{I}_K + \frac{P_{\mathrm{t}}}{\sigma_z^2}\frac{\pi}{4}\mathbf{P}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}\mathbf{F}_{\mathrm{d}}\mathbf{F}_{\mathrm{d}}^{\mathrm{H}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}}\Big) \\ &= C_{\mathrm{sum}}(\frac{\pi}{4}P_{\mathrm{t}}, \mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{F}_{\mathrm{d}}) = C_{\mathrm{sum}}(\frac{\pi}{4}P_{\mathrm{t}}, \frac{1}{\sqrt{\Gamma_{\mathrm{t}}}}\mathbf{H}_{\mathrm{e}}\mathbf{F}_{\mathrm{ZFe}}) \\ &= R_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{\Gamma_{\mathrm{t}}}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}\mathbf{F}_{\mathrm{ZFe}}) = K\log_2(1+\frac{\pi P_{\mathrm{t}}}{4K\Gamma_{\mathrm{t}}\sigma_z^2}). \end{aligned} \tag{46}$$

Hence, by letting the baseband precoder for the hybrid beamformer with constant modulus constraint as $\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}} = \mathbf{F}_{\mathrm{ZFe}}$ combined with $\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}$ of Lemma 2, the asymptotically optimal hybrid beamformer is achieved. Finally, it could be easily verified that

$$C_{\mathrm{sum}}(P_{\mathrm{t}}, \mathbf{H}) - R_{\mathrm{sum}}(P_{\mathrm{t}}, \frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{H}\mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}\mathbf{F}_{\mathrm{B}}^{\mathrm{opt}}) = \lim_{N_{\mathrm{t}}\to\infty} K\log_2\frac{1+\frac{P_{\mathrm{t}}}{K\Gamma_{\mathrm{t}}\sigma_z^2}}{1+\frac{\pi P_{\mathrm{t}}}{4K\Gamma_{\mathrm{t}}\sigma_z^2}} = -K\log_2(\pi/4) \tag{47}$$

APPENDIX F
PROOF OF LEMMA 6

According to Appendix A, the spectral efficiency achieved by the hybrid beamformer depends on $\frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{v}_k^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k}$. When the RF beamformer is set based on (22),

$$\lim_{N_{\mathrm{t}}\to\infty}\frac{1}{\sqrt{N_{\mathrm{t}}}}\mathbf{v}_k^{\mathrm{H}}\mathbf{f}_{\mathrm{RF},k} = \lim_{N_{\mathrm{t}}\to\infty}\frac{1}{\sqrt{N_{\mathrm{t}}}}\sum_{n_{\mathrm{t}}=1}^{N_{\mathrm{t}}} V_{n_{\mathrm{t}}k}^* F_{\mathrm{RF},n_{\mathrm{t}}k} = \lim_{N_{\mathrm{t}}\to\infty}\frac{1}{N_{\mathrm{t}}}\sum_{n_{\mathrm{t}}=1}^{N_{\mathrm{t}}}\sqrt{N_{\mathrm{t}}}V_{n_{\mathrm{t}}k}^* F_{\mathrm{RF},n_{\mathrm{t}}k} = \mathrm{E}[\tilde{V}_{n_{\mathrm{t}}k}], \tag{48}$$

where $\tilde{V}_{n_{\mathrm{t}}k}$ is defined as

$$\tilde{V}_{n_{\mathrm{t}}k} = \begin{cases} 0, & \sqrt{N_{\mathrm{t}}}|V_{n_{\mathrm{t}}k}| \leq \alpha, \\ \sqrt{N_{\mathrm{t}}}|V_{n_{\mathrm{t}}k}|, & \alpha < \sqrt{N_{\mathrm{t}}}|V_{n_{\mathrm{t}}k}|. \end{cases} \tag{49}$$

Theorem 1 states that $\sqrt{N_{\mathrm{t}}}|V_{n_{\mathrm{t}}k}|$ follows a Rayleigh distribution with parameter σ_{R}. As a result, the PDF of $\tilde{V}_{n_{\mathrm{t}}k}$ is expressed as

$$\Pr(\tilde{V}) = \begin{cases} \Pr(\sqrt{N_{\mathrm{t}}}|V| \leq \alpha)\delta(0), & \tilde{V} \leq \alpha, \\ \frac{\tilde{V}}{\sigma_{\mathrm{R}}^2}\mathrm{e}^{-\tilde{V}^2/2\sigma_{\mathrm{R}}^2}, & \alpha < \tilde{V}. \end{cases} \tag{50}$$

The expected value of $\tilde{V}$ is calculated as

$$\mathrm{E}[\tilde{V}(\alpha)] = \int_{-\infty}^{+\infty}\tilde{V}\Pr(\tilde{V})d\tilde{V} = \int_{\alpha}^{+\infty}\frac{\tilde{V}^2}{\sigma_{\mathrm{R}}^2}\mathrm{e}^{-\tilde{V}^2/2\sigma_{\mathrm{R}}^2}d\tilde{V} \tag{51}$$

$$= \int_0^{+\infty} \frac{\tilde{V}^2}{\sigma_{\mathrm{R}}^2} \mathrm{e}^{-\tilde{V}^2/2\sigma_{\mathrm{R}}^2} d\tilde{V} - \int_0^{\alpha} \frac{\tilde{V}^2}{\sigma_{\mathrm{R}}^2} \mathrm{e}^{-\tilde{V}^2/2\sigma_{\mathrm{R}}^2} d\tilde{V} \overset{(g)}{=} \sigma_{\mathrm{R}} \sqrt{\frac{\pi}{2}} - \int_0^{\alpha} \frac{\tilde{V}^2}{\sigma_{\mathrm{R}}^2} \mathrm{e}^{-\tilde{V}^2/2\sigma_{\mathrm{R}}^2} d\tilde{V}$$
$$\overset{(h)}{=} \sigma_{\mathrm{R}} \sqrt{\frac{\pi}{2}} - \sqrt{2}\sigma_{\mathrm{R}} \Big(\frac{\sqrt{\pi}}{2} \mathrm{erf}(\frac{\alpha}{\sqrt{2}\sigma_{\mathrm{R}}}) - \frac{\alpha}{\sqrt{2}\sigma_{\mathrm{R}}} \mathrm{e}^{-\alpha^2/2\sigma_{\mathrm{R}}^2} \Big) = \frac{\sqrt{\pi}}{2} + \alpha \mathrm{e}^{-\alpha^2} - \frac{\sqrt{\pi}}{2} \mathrm{erf}(\alpha),$$

where (g) and (h) are derived from [29]. Moreover, the cumulative distribution function (CDF) of $\sqrt{N_{\mathrm{t}}}|V| \leq \alpha$ is expressed as $\Pr(|V_{n_{\mathrm{t}}k}| \leq \alpha) = \beta/100 = 1 - \mathrm{e}^{-\alpha^2/2\sigma_{\mathrm{R}}^2}$, hence $\alpha = \sqrt{-\ln(1-\beta/100)}$ where β is the percentage of the phase shifters that are turned off. It could be easily shown that $\frac{1}{\sqrt{N_{\mathrm{t}}}} |\mathbf{v}_k^{\mathrm{H}} \mathbf{f}_{\mathrm{RF},k'}| = 0$, $\forall k \neq k'$ and $\frac{1}{\sqrt{N_{\mathrm{t}}}} \mathbf{F}_{\mathrm{d}}^{\mathrm{H}} \mathbf{F}_{\mathrm{RF}}$ becomes a diagonal matrix with equal diagonal elements. Hence, the baseband precoder matrix becomes $\mathbf{F}_{\mathrm{B}} = \mathbf{I}_K$. Applying the same phase shifter selection scheme at the receiver side, it can be easily verified that $\Gamma_{\mathrm{t}} = (1-\beta/100)N_{\mathrm{t}}$, $\Gamma_{\mathrm{r}} = (1-\beta/100)N_{\mathrm{r}}$, $\mathbf{R}_{\mathrm{n}} = 1/\Gamma_{\mathrm{r}} \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{W}_{\mathrm{RF}}^{\mathrm{H}} \mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{B}} = \mathbf{I}_K$ and $1/\sqrt{N_{\mathrm{t}}} \mathbf{F}_{\mathrm{d}}^{\mathrm{H}} \mathbf{F}_{\mathrm{RF}} = \mathrm{E}[\tilde{V}(\alpha)] \mathbf{I}_K$. Similar to Appendix C, the spectral efficiency is expressed as

$$R_\beta = \lim_{N_{\mathrm{t}} \to \infty} \lim_{N_{\mathrm{r}} \to \infty} \log_2 \det \Big(\frac{\rho}{(1-\beta)^2 N_{\mathrm{t}} N_{\mathrm{r}}} \mathbf{R}_{\mathrm{n}}^{-1} \mathbf{W}_{\mathrm{B}}^{\mathrm{H}} \mathbf{W}_{\mathrm{RF}}^{\mathrm{H}} \mathbf{H} \mathbf{F}_{\mathrm{RF}} \mathbf{F}_{\mathrm{B}} \mathbf{P} \mathbf{F}_{\mathrm{B}}^{\mathrm{H}} \mathbf{F}_{\mathrm{RF}}^{\mathrm{H}} \mathbf{H}^{\mathrm{H}} \mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{B}} \Big) \tag{52}$$
$$= \lim_{N_{\mathrm{t}} \to \infty} \lim_{N_{\mathrm{r}} \to \infty} \log_2 \det \Big(\frac{\rho}{(1-\beta)^2 N_{\mathrm{t}} N_{\mathrm{r}}} \mathbf{W}_{\mathrm{RF}}^{\mathrm{H}} \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\mathrm{H}} \mathbf{F}_{\mathrm{RF}} \mathbf{P} \mathbf{F}_{\mathrm{RF}}^{\mathrm{H}} \mathbf{V} \mathbf{\Sigma} \mathbf{U}^{\mathrm{H}} \mathbf{W}_{\mathrm{RF}} \Big)$$
$$= \sum_{k=1}^{K} \log_2 \left(\frac{\rho P_{kk} \sigma_{kk}^2}{(1-\beta)^2} \Big(\mathrm{E}[\tilde{V}(\alpha)] \Big)^4 \right)$$
$$\overset{(i)}{=} \sum_{k=1}^{K} \log_2 (\rho P_{kk} \sigma_{kk}^2) + 4K \log_2 \Big(\mathrm{E}[\tilde{V}(\alpha)] \Big) - 2K \log_2 (1-\beta),$$

as the first term after (i) is equal to C at high SNR. □

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°619563 (MiWaveS). We would also like to acknowledge the support of the University of Surrey 5GIC members for this work.

REFERENCES

[1] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 40–60, January 2013.

[2] L. Godara, "Application of antenna arrays to mobile communications. ii. beam-forming and direction-of-arrival considerations," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1195–1245, August 1997.

[3] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *Communications Magazine, IEEE*, vol. 52, no. 12, pp. 122–131, December 2014.

[4] P. Sudarshan, N. Mehta, A. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3501–3511, December 2006.

[5] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing - the large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, October 2013.

[6] A. Adhikary, E. Al Safadi, M. Samimi, R. Wang, G. Caire, T. Rappaport, and A. Molisch, "Joint spatial division and multiplexing for mm-Wave channels," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1239–1255, June 2014.

[7] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4505–4515, September 2014.

[8] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, March 2014.

[9] S. Payami, M. Shariat, M. Ghoraishi, and M. Dianati, "Effective RF codebook design and channel estimation for millimeter wave communication systems," *IEEE International Conference on Communication (ICC) Workshops*, pp. 1226–1231, June 2015.

[10] W. Ni, X. Dong, and W. Lu, "Near-optimal hybrid processing for massive MIMO systems via matrix decomposition," *CoRR*, vol. abs/1504.03777, 2015. [Online]. Available: http://arxiv.org/abs/1504.03777

[11] X. Zhang, A. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4091–4103, November 2005.

[12] A. Alkhateeb, G. Leus, and R. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2015.

[13] Z. Pi, "Optimal transmitter beamforming with per-antenna power constraints," *IEEE International Conference on Communications (ICC)*, pp. 3779–3784, June 2012.

[14] S. Sohrabi, , and Y. Wei, "Hybrid digital and analog beamforming design for large-scale MIMO systems," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015.

[15] ——, "Hybrid beamforming with finite-resolution phase shifters for large-scale MIMO systems," *IEEE International Workshop on Signal Processing for Wireless Communications (SPAWC)*, July 2015.

[16] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press, 2005.

[17] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, June 2003.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[19] D. Love, R. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE International Conference on Communications (ICC)*, vol. 4, pp. 2618–2622 vol.4, May 2003.

[20] M. Spruill, "Asymptotic distribution of coordinates on high dimensional spheres," *Electron. Commun. Probab.*, vol. 12, pp. no. 23, 234–247, 2007.

[21] C. Walck, *Hand-book on statistical distributions for experimentalists*, December 1996.

[22] O. El Ayach, R. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter

wave MIMO systems," *IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 100–104, June 2012.

[23] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and its Applications*, vol. 415, no. 1, pp. 20–30, 2006.

[24] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, August 2003.

[25] Y. Zheng and C. E. Saavedra, "An ultra-compact CMOS variable phase shifter for 2.4-GHz ISM applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 6, pp. 1349–1354, June 2008.

[26] N. A. Talwalkar, C. P. Yue, H. Gan, and S. S. Wong, "Integrated CMOS transmit-receive switch using LC-tuned substrate bias for 2.4-GHz and 5.2-GHz applications," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 6, pp. 863–870, June 2004.

[27] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[28] N. Jindal, "High SNR analysis of MIMO broadcast channels," *Proceedings of International Symposium on Information Theory (ISIT)*, pp. 2310–2314, September 2005.

[29] I. S. Gradshteyn and I. M. Ryzhik, "Table of integrals, series, and products," *Elsevier Academic Press, Amsterdam*, 2007.