Energy Efficient Hybrid Beamforming in Massive MU-MIMO Systems via Eigenmode Selection

Weiheng Ni, Po-Han Chiang, and Sujit Dey
Mobile Systems Design Lab, Dept. of Electrical and Computer Engineering, University of California, San Diego
La Jolla, CA, USA
{whni, pochiang}@ucsd.edu, dey@ece.ucsd.edu

***Abstract*—Hybrid beamforming in massive multiple-input multiple-output (MIMO) systems is one of the most promising techniques to meet rapidly increasing data rate demand of the fifth generation cellular system with reduced implementation costs. In this paper, we focus on a downlink communication scenario where a hybrid beamforming base station (BS) transmits data to multiple single-antenna users. Given the knowledge of channel state information of all users, the hybrid beamformers at the BS are designed to minimize the BS power consumption while the data rate needed to meet quality of service (QoS) requirement of each user is satisfied. Herein, the zero-forcing (ZF) beamforming is directly applied on the effective baseband channel, and the radio frequency (RF) beamformer is generated by matching the beamforming matrix columns, selected from a preset discrete Fourier transform (DFT) basis codebook, with the eigenvectors of the aggregated propagation channel, which is termed as eigenmode selection. We also present a phase array batch-switching structure to realize the eigenmode selection beamforming economically. Simulations demonstrate that substantial transmission power can be saved with the proposed eigenmode selection beamforming compared to existing propagation path matching schemes, especially in rich-scattering channels, while satisfying given QoS (data rate) requirements.**

***Index Terms*—Massive MIMO, hybrid architecture, multiple users, transmission power, QoS requirement, eigenmode selection.**

I. INTRODUCTION

The ever-growing mobile data traffic has been pushing wireless communications toward the fifth generation (5G) standards, which promises a roughly 1000-fold increase on area spectral efficiency. As one of the key technologies in 5G cellular systems, massive multi-input multiple-output (MIMO) highlights its importance by enormously elevating the spectral efficiency [1]. However, other than rate improvement, the traditional architecture for massive MIMO is not energy efficient. As the number of antennas goes to a very large scale, the implementation complexity and energy consumption of the massive MIMO systems will become economically unacceptable [2]. The main reason is that each antenna is coupled with a dedicated radio frequency (RF) chain, the hardware set that includes digital-to-analog (D/A) converter, mixer, and power amplifier (PA). A potential solution to this problem origins from hybrid beamforming architectures, enabling the use of a limited number of RF chains in massive MIMO systems [3]. Instead of applying only the full-complexity baseband digital processing, the hybrid transceiver employs a low-dimensional baseband beamformer, cascaded with an analog beamformer that consists of a large quantity of phase shifters in the RF domain, which provide phase-only control on the processed signals. In other words, hybrid beamforming is highly promising in reaching a balance between the rate enhancement and power saving.

The hybrid beamforming architectures currently studied in the literature can be classified into two main categories: i) fully-connected and ii) array-of-subarray structures. The former has a potential to approach the spectral efficiency achieved by full-complexity beamforming while the latter takes advantages of reducing the number of phase shifters but with the cost of performance degradation [4]. The effect of the RF beamformer on the processed signals can be characterized as the multiplication with a matrix with only equal-amplitude entries and some zeros, depending on which architecture is chosen. Therefore, in hybrid beamforming design, one intuitive way is to approximate the full-complexity beamformer matrix with the products of hybrid (digital and RF) beamformers, especially in the single-user MIMO (SU-MIMO) scenario where the optimal full-complexity beamformer can be found [3]. Some techniques, e.g., alternate matrix decomposition and orthogonal matching pursuit, are proposed to realize such kind of approximation in rich/poor-scattering environments [5], [6]. In addition, when it comes to multi-user MIMO (MU-MIMO) systems, decoupling the design of the RF and baseband beamformers is more acceptable due to the reduced complexity, where RF beamforming aims to harvest the channel gains while the baseband beamforming works on the effective baseband (RF chain to RF chain) channels [2], [7].

Besides optimizing the spectral efficiency of massive MIMO systems, power consumption has also generated considerable interest in terms of operation costs for mobile operators [8]. Recent analysis of energy efficiency (spectral efficiency divided by power consumption) in traditional full-complexity MIMO systems have indicated that energy efficiency is not always enhanced along with an increasing number of antennas, while the transmission power keeps decreasing [9], [10], [11], [12]. This fact facilitates the migration to hybrid beamforming systems, since it could possibly reduce the number of RF chains to the quantity of supported data streams and save substantial power consumed by those RF chains. [12] elaborates the power consumption models for both fully-connected and array-of-subarray hybrid beamforming architectures, and demonstrates their capability to achieve higher energy efficiency rather than

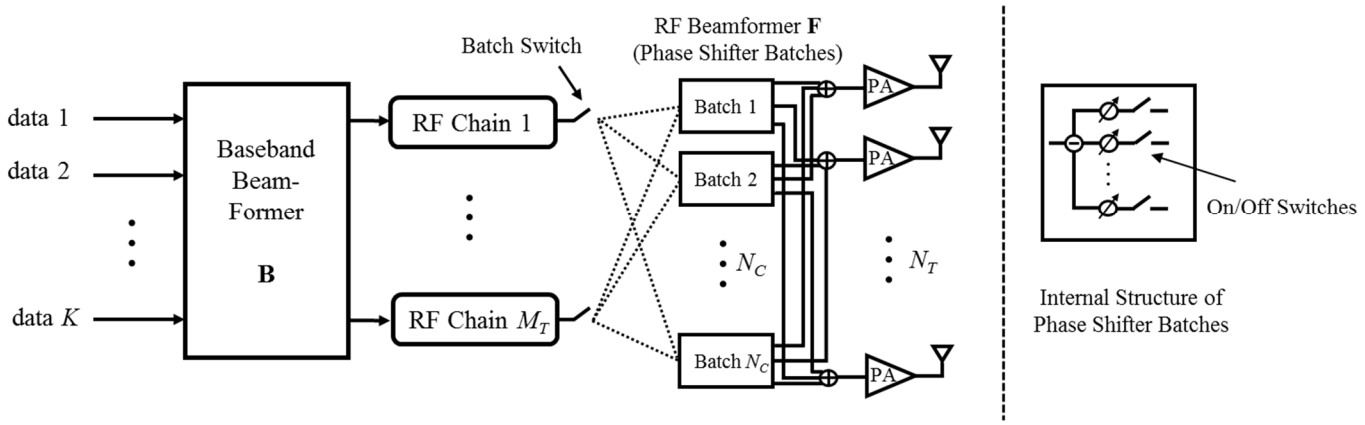

Fig. 1. System diagram of a fully-connected hybrid beamforming BS implemented with phase shifter batches.

the full-complexity systems. So far, most research efforts have focused on the fundamental limits on the energy efficiency rather than optimizing power consumption under practical application scenarios.

In this paper, we focus on a downlink communication scenario where a hybrid beamforming base station (BS) transmits data to multiple single-antenna users. It is assumed that the instantaneous transmission data rate, referred to as the quality of service (QoS), requirements of all users are satisfied. Given the knowledge of channel state information of all users, we propose a fully-connected hybrid beamforming design to minimize the BS (transmission) power consumption when considering the QoS (data rate) requirements. This power minimization objective inspires the hybrid beamforming design from a perspective that is different from those common rate-optimization methods. For simplicity, the zero-forcing (ZF) beamforming is directly applied on the effective baseband channel, while the RF beamformer is generated by matching the beamforming matrix columns, selected from a preset discrete Fourier transform (DFT) basis codebook, with the eigenvectors of the aggregated propagation channel, which is termed as eigenmode selection. Inspired by [13], we present a phase array batch-switching structure to realize the eigenmode selection beamforming economically, where each RF chain selects a batch of fixed phase shifters corresponding to a vector in the codebook instead of using variable phase shifters. The contributions of this paper can be summarized as follows:

- This paper offers the first approach to design hybrid beamformers with the goal of minimizing power while satisfying given users' QoS requirements;
- The low-complexity eigenmode selection beamforming scheme works effectively with all types of channels;
- The phase array batch-switching structure gives an alternative to economically realize the fully-connected hybrid beamformer structure.

Simulation results demonstrate that less BS total/transmission power is consumed in average with the proposed eigenmode selection beamforming, without compromising QoS (data rate) required, compared to existing propagation path matching schemes, e.g., in [12], especially in rich-scattering channels. In addition, the outage probability performance is also examined in the simulations.

The rest of paper is organized as follows. Section II elaborates the fully-connected hybrid beamforming structure and introduces the involved channel and BS power models. Section III presents the derivations of the eigenmode selection scheme and the corresponding algorithm. The performance comparison is conducted in Section IV. We finally conclude this work in Section V.

Notation: $(\cdot)^T$, $(\cdot)^H$and $(\cdot)^{-1}$denote the transpose, conjugate transpose, and inverse respectively. $\mathbb{E}[\cdot]$ represents expectation over random variables. $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$ indicates the 2-norm operation over a vector and the Frobenius norm over a matrix. $\mathrm{Tr}(\cdot)$ denotes the trace of a square matrix.

II. System Description

In this section, we give a detailed description of the fully-connected hybrid beamforming system, the involved channel model, and the BS power consumption model. Note that we decide to consider the fully-connected structure, instead of the array-of-subarray structure, due to the former's potential to approach the spectral efficiency achieved by full-complexity beamforming.

A. System Model

We consider an MU-MIMO downlink communication with K single-antenna users. As shown in Fig. 1, the fully-connected structure is applied for the hybrid beamforming BS, which is equipped with M_T RF chains and N_T transmit antennas. To guarantee the effectiveness of spatial multiplexing, the number of data streams (one stream per single-antenna user) is constrained by $K \le M_T \le N_T$. The transmit symbols at first are processed by the baseband digital beamformer **B** of dimension $M_T \times K$ and then by an analog RF beamformer **F** of dimension $N_T \times M_T$ after frequency up-conversion at M_T RF chains. Note that the baseband beamformer **B** is capable of modifying signal amplitude and phase while only phase changes can be realized by the RF beamformer since it is implemented with an array of phase shifters. In the fully-connected architecture, all entries in RF beamformer **F** are with equal-amplitude. Each entry in **F** is normalized as $|\mathbf{F}^{(i,j)}|=\frac{1}{\sqrt{N_T}}$, where $\mathbf{F}^{(i,j)}$ is the (i, j)-th entry of **F**.

In this paper, we assume that all the columns of $\mathbf{F}$ are selected from a preset codebook $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_{N_C}\}$, where each basis vector $\mathbf{d}_n$, n=1, ⋯, N_C, in length of N_T, also satisfies the equal-amplitude constraint. Normally, the number of bases in codebook $\mathcal{D}$ does not exceed the vector dimension, which is $N_C \leq N_T$. In this sense, we propose a novel phase shifter batch-switching structure that implements the fully-connected RF beamformer in Fig. 1. In the codebook $\mathcal{D}$, the effect of a basis vector $\mathbf{d}_n$ can be realized with a batch of N_T phase shifters that are tuned to the phases of all elements in $\mathbf{d}_n$ and connected to all N_T antennas, as shown at the right side of Fig. 1. Consequently, if a design requires to select M_T bases from codebook $\mathcal{D}$ to form the $N_T \times M_T$ RF beamformer, the action performed on such a batch-switching structure is just switching all RF chains to M_T distinct phase shifter batches that are corresponding to the target basis vectors, illustrated in the left side of Fig. 1. All selected phase shifter batches are cascaded in circuits via closing their internal switches, while the rest of batches open their internal switches. The benefit brought by this phase shifter batch-switching structure is that we do not tune any phase shifter to realize beamforming except performing accurate switching, which is quite suitable to the codebook selection based design. Nevertheless, we must carefully choose the size and basis vectors of the codebook during design since $N_T \times N_C$ phase shifters are required and tuning operation is not allowed when performing beamforming. In the proposed hybrid beamforming design, we choose N_T-dimensional complex DFT basis set as the codebook, based on the observation of the channel model we investigate in Section II-B.

Other than the difference in terms of implementation, the analytical model is identical to the normal fully-connected hybrid beamforming architecture. We assume a narrowband flat fading channel and obtain the received signals at all users

$$y_k = \mathbf{h}_k \mathbf{F} \mathbf{B} \mathbf{s} + n_k, \quad k = 1, 2, \cdots, K, \tag{1}$$

where $\mathbf{s} = [s_1, s_2, \cdots, s_K]^T \in \mathbb{C}^{K \times 1}$ is the transmit vector for a total of K users and s_k is the data symbol of the k-th single-antenna user, which satisfies $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \frac{P_s}{K}\mathbf{I}_K$. Here, $\mathbf{I}_K$ denotes a $K \times K$ identity matrix and P_s is the initial transmission power. The channel vector for the k-th user $\mathbf{h}_k$ is of dimension $1 \times N_T$, and all n_k's are i.i.d $\mathbb{CN}(0, \sigma^2)$ complex Gaussian noise. We can also aggregate all received signals in a vector as follows

$$\mathbf{y} = \mathbf{H}\mathbf{F}\mathbf{B}\mathbf{s} + \mathbf{n}, \tag{2}$$

where $\mathbf{y} = [y_1, y_2, \cdots, y_K]^T$, $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_K]^T$ and $\mathbf{n} = [n_1, n_2, \cdots, n_K]^T$. The channel state information $\mathbf{H}$ is assumed to be known at the BS, and the effective baseband channel is defined as $\mathbf{H}_{eff} = \mathbf{H}\mathbf{F}$.

B. Channel Model

We adopt a geometric channel model [2], [4], which considers the summing effect of multiple propagation paths. A general channel notation for the k-th user is as $\mathbf{h}_k = \beta_k \tilde{\mathbf{h}}_k$, where β_k indicates the large-scale path fading and $\tilde{\mathbf{h}}_k$ represents the normalized channel vector, satisfying that $\mathbb{E}[||\tilde{\mathbf{h}}_k||_2^2] = N_T$. The multi-path summation effect is described as

$$\tilde{\mathbf{h}}_k = \sqrt{\frac{N_T}{N_P}} \sum_{p=1}^{N_P} \alpha_p^k \mathbf{a}^H(\phi_p^k, \varphi_p^k), \tag{3}$$

where N_P propagation paths are assumed to be observed by each user. The complex channel gain of the p-th path α_p^k follows $\mathbb{CN}(0, 1)$. $\mathbf{a}^H(\phi_p^k, \varphi_p^k)$ is the array response vector at angles of departure (AoDs) ϕ_p^k and φ_p^k, corresponding to the azimuth and elevation directions. As for the statistics of (ϕ_p^k, φ_p^k), i.i.d uniform distribution is enforced on all paths for one user. The lower and upper bound for (ϕ_p^k, φ_p^k) are defined as $([\phi_{\min}^k, \phi_{\max}^k], [\varphi_{\min}^k, \varphi_{\max}^k])$. Specially, when $([\phi_{\min}^k, \phi_{\max}^k], [\varphi_{\min}^k, \varphi_{\max}^k])$ are chosen as $([-\pi, \pi], [-\pi, \pi])$ and there are an unlimited number of paths, $N_P \to \infty$, we reach the case of Rayleigh fading.

The BS employs a uniform linear array (ULA), while the proposed eigenmode selection scheme is not limited to it. For an N_T-element ULA antenna array, only AoDs in azimuth direction are considered, and the array response vector is given by

$$\mathbf{a}_{ULA}(\phi) = \sqrt{\frac{1}{N_T}} \left[1, e^{j\frac{2\pi}{\lambda} d \sin\phi}, \cdots, e^{j(N_T-1)\frac{2\pi}{\lambda} d \sin\phi}\right]^T,$$

where λ is the carrier wavelength, and d is the antenna spacing, typically equal to half of λ. It is demonstrated that the beam direction matching is effective for RF beamformer design [5]. Based on this fact, the preset codebook can be selected to approximate the above ULA array response vector. Note that N_T-dimensional DFT basis has the same form $\mathbf{d}(\omega) = \sqrt{1/N_T}\,[1, e^{j\omega}, \cdots, e^{j(N_T-1)\omega}]^T$ given $\omega = \frac{2\pi}{\lambda} d \sin\phi$ [7]. Therefore, the beamforming codebook $\mathcal{D}_{\text{DFT}} = \{\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_{N_C}\}$ can be set to a full DFT basis set with $\mathbf{d}_n = \mathbf{d}(\frac{2\pi}{N_T} n)$ and $N_C = N_T$.

C. Power Consumption Model

The primary interest in this paper is to minimize the BS power consumption through beamforming design while the QoS requirements of users are satisfied. Hence, it is of importance to evaluate the total BS power including the power consumption of circuitry, with a realistic model. A common BS power model mainly considers the consumption of baseband processing P_{BP}, RF chains P_{RF}, and PAs P_{PA} [14], calculated as

$$P_{tot} = \frac{P_{BP} + P_{RF} + P_{PA}}{(1 - \sigma_{DC})(1 - \sigma_{MS})(1 - \sigma_{cool})}, \tag{4}$$

where σ_{DC}, σ_{MS} and σ_{cool} are the loss factors resulting from direct current (DC) conversion, main supply (MS) and cooling system. The power consumption of P_{PA} can be estimated with

$$P_{PA} = \frac{P_{tx}}{\eta_{PA}(1 - \sigma_{feed})} = \frac{\sqrt{P_{tx} P_{tx}^{\max}}}{\eta_{PA}^{\max}(1 - \sigma_{feed})},$$

where the PA efficiency (class-B) $\eta_{PA} = \eta_{PA}^{\max}\sqrt{P_{tx}/P_{tx}^{\max}}$, and σ_{feed}, $\eta_{PA}^{\max}$, P_{tx} ($P_{tx}^{\max}$) represent the feeder loss, the peak PA efficiency, and the output transmission power (its maximum) at antennas [12], [14]. Typically, P_{BP} relies on the dimension of baseband processing and P_{RF} is proportional to the number of

active RF chains M_T, which is $P_{RF} = M_T P_{RF,0}$, where $P_{RF,0}$ is the circuit power per RF chain. Note that only the transmission power P_{tx} is affected by the hybrid beamforming schemes. As for other components contained in (4), they are regarded to be fixed for different beamforming designs.

III. Eigenmode Selection Hybrid Beamforming

To conduct the hybrid beamforming with an objective of BS power minimization, we choose to decouple the design of RF and baseband beamformers. Given a ZF baseband beamformer **B**, an eigenmode selection RF beamforming scheme is proposed to pursue less power use. The theoretical derivations and the corresponding algorithm will be presented in this section.

A. Problem Formulation with Baseband ZF Beamformer

Given the baseband effective channel $\mathbf{H}_{eff} = \mathbf{HF}$, ZF is adopted for the baseband ZF beamformer to cancel inter-user interference, which is implemented through pseudo-inverse as

$$\mathbf{B} = \mathbf{H}_{eff}^H (\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1}\mathbf{\Lambda}, \tag{5}$$

where $\mathbf{\Lambda} = \mathrm{diag}\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_K}\}$, and λ_k is the normalized factor to regulate the receive signal-to-noise (SNR) at the k-th user [15]. With such a baseband ZF beamformer, the aggregated received signals can be updated as

$$\begin{aligned}\mathbf{y} &= \mathbf{HFBs} + \mathbf{n}\\ &= (\mathbf{HF})\mathbf{H}_{eff}^H (\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1}\mathbf{\Lambda s} + \mathbf{n}\\ &= [\sqrt{\lambda_1}s_1, \sqrt{\lambda_2}s_2, \cdots, \sqrt{\lambda_K}s_K]^T + \mathbf{n}.\end{aligned}$$

The achievable rate for the k-th user is

$$\begin{aligned}r_k &= W\log(1+\mathrm{SNR}_k)\\ &= W\log\left(1+\frac{P_s\lambda_k}{K\sigma^2}\right),\ k = 1, \cdots, K.\end{aligned} \tag{6}$$

where W is the system bandwidth. To realize the achievable rates for all users, the transmission power at BS antennas is given by

$$P_{tx} = \|\mathbf{FBs}\|_2^2 = \frac{P_S}{K}\sum_{k=1}^K \lambda_k \left[\ (\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1}\right]_{k,k}, \tag{7}$$

where $[\,\cdot\,]_{k,k}$ operator takes the k-th diagonal entry of a square matrix. Each term of the summation in (7) can be regarded as the power consumption used to communicate with each user, which depends on λ_k, an SNR factor directly relying on the achievable rate according to (6), and $\left[\ (\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1}\right]_{k,k}$, which is positively affected by the channel gains for the k-th user and negatively influenced by the interference from other users. Given the QoS requirement (minimum rate) set of all users, $\{\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_K\}$, the BS power problem is as follows,

$$\begin{aligned}&\min_{\mathbf{B},\mathbf{F}}\ \{P_{tot}\}\\ &\text{s.t. } W\log\left(1+\frac{P_s\lambda_k}{K\sigma^2}\right) \ge \bar{r}_k, \text{ for all } k,\\ &\mathbf{B} = \mathbf{H}_{eff}^H (\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1}\mathbf{\Lambda},\\ &\mathbf{f}_m \in \mathcal{D}_{\mathrm{DFT}},\ m = 1, \cdots, M_T,\end{aligned} \tag{8}$$

where $\mathbf{f}_m$ is the m-th column of RF beamformer matrix **F**. Note that all terms of P_{tot} in (4), except P_{tx}, can be regarded as constants with respect to the variables **B** and **F**. Hence, $\min_{\mathbf{B},\mathbf{F}}\{P_{tot}\} \Leftrightarrow \min_{\mathbf{B},\mathbf{F}}\{\sqrt{P_{tx}}\} \Leftrightarrow \min_{\mathbf{B},\mathbf{F}}\{P_{tx}\}$, which means

$$\min_{\mathbf{B},\mathbf{F}}\ \{P_{tot}\} \Leftrightarrow \min_{\mathbf{B},\mathbf{F}}\left\{\sum_{k=1}^K \lambda_k \left[\ (\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1}\right]_{k,k}\right\}. \tag{9}$$

In addition, the constraints rendered on **B** and **F** are separable in (8), so that we are allowed to solve the power minimization problem with respect to the baseband beamformer **B** first, assuming **F** is known.

From this perspective, only $\mathbf{\Lambda} = \mathrm{diag}\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_K}\}$ in **B** has not been determined. With the observation of (8) and (9), it is evident that the optimal $\mathbf{\Lambda}$ can be obtained when all λ_k's reach their own minimums, given by

$$\lambda_k^* = \frac{K\sigma^2}{P_s}\left(2^{\bar{r}_k/W} - 1\right),\ k = 1, \cdots, K. \tag{10}$$

As long as the RF beamformer **F** is provided, the baseband beamformer can be obtained based on (5). Note that, the final transmission power does not depend on the initial symbol power P_s since it can be eliminated by the product of $P_s\lambda_k$ in (7).

B. Eigenmode Selection to Realize RF Beamforming

To further explore the essence of RF beamforming design, we need to make some transforms on the objective function as

$$\begin{aligned}&\sum_{k=1}^K \lambda_k \left[\ (\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1}\right]_{k,k}\\ &= \mathrm{Tr}\left[\mathbf{\Lambda}\ (\mathbf{HFF}^H\mathbf{H}^H)^{-1}\right]\\ &= \mathrm{Tr}\left[\ (\widehat{\mathbf{H}}\mathbf{FF}^H\widehat{\mathbf{H}}^H)^{-1}\right],\end{aligned} \tag{11}$$

where $\widehat{\mathbf{H}} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{H}$ contains the QoS requirements and channel state information of all users. Since $K \le M_T$ always holds for the RF beamformer **F**, we attempt to divide M_T-column **F** into sub-matrices with K columns in the case that M_T is a multiple of K. Then we have

$$\mathbf{F} = [\mathbf{F}^{(1)}, \cdots, \mathbf{F}^{(b)}, \cdots, \mathbf{F}^{(M_T/K)}],\ b = 1, \ldots, M_T/K,$$

where $\mathbf{F}^{(b)}$ is a matrix that contains elements from the $[b(K-1)+1]$-th column to the (bK)-th column of **F**. (11) can be further decomposed as

$$\begin{aligned}\mathrm{Tr}\left[\ (\widehat{\mathbf{H}}\mathbf{FF}^H\widehat{\mathbf{H}}^H)^{-1}\right] &= \mathrm{Tr}\left[\left(\widehat{\mathbf{H}}\sum_b \mathbf{F}^{(b)}\mathbf{F}^{(b)H}\widehat{\mathbf{H}}^H\right)^{-1}\right]\\ &= \sum_b \mathrm{Tr}\left[\left(\widehat{\mathbf{H}}\mathbf{F}^{(b)}\mathbf{F}^{(b)H}\widehat{\mathbf{H}}^H\right)^{-1}\right].\end{aligned} \tag{12}$$

So far, the joint design of all $\mathbf{F}^{(b)}$'s is still intractable since the columns of all different $\mathbf{F}^{(b)}$'s are not equal to each other, such that all trace terms in (12) are not independent. To simplify the optimization procedures, we choose to determine $\mathbf{F}^{(b)}$'s in the order of $b = 1, 2, \ldots$, via solving a sub-problem of minimizing $\mathrm{Tr}\left[\left(\widehat{\mathbf{H}}\mathbf{F}^{(b)}\mathbf{F}^{(b)H}\widehat{\mathbf{H}}^H\right)^{-1}\right]$. The insight behind $\widehat{\mathbf{H}}\mathbf{F}^{(b)}\mathbf{F}^{(b)H}\widehat{\mathbf{H}}^H$ is

that the channel gain of $\widehat{\mathbf{H}}$ can be harvested by the columns of the b-th RF beamformer block $\mathbf{F}^{(b)}$, making the diagonal entries of $\widehat{\mathbf{H}}\mathbf{F}^{(b)}$ as large as possible, so as to reduce the BS power use which is represented by the trace terms. Another rationality to decompose $\mathbf{F}$ lies in the fact that $\widehat{\mathbf{H}}\mathbf{F}^{(b)}$ is an invertible square matrix that enables the exchange between $\widehat{\mathbf{H}}\mathbf{F}^{(b)}$ and $\mathbf{F}^{(b)^H}\widehat{\mathbf{H}}^H$ in the trace operator, which is

$$\begin{aligned}
&\mathrm{Tr}\left[\left(\widehat{\mathbf{H}}\mathbf{F}^{(b)}\mathbf{F}^{(b)^H}\widehat{\mathbf{H}}^H\right)^{-1}\right] \\
&\overset{(a)}{=} \mathrm{Tr}\left[\left(\mathbf{F}^{(b)^H}\widehat{\mathbf{H}}^H\right)^{-1}\left(\widehat{\mathbf{H}}\mathbf{F}^{(b)}\right)^{-1}\right] \\
&\overset{(b)}{=} \mathrm{Tr}\left[\left(\widehat{\mathbf{H}}\mathbf{F}^{(b)}\right)^{-1}\left(\mathbf{F}^{(b)^H}\widehat{\mathbf{H}}^H\right)^{-1}\right] \\
&= \mathrm{Tr}\left\{\left[\mathbf{F}^{(b)^H}(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}})\mathbf{F}^{(b)}\right]^{-1}\right\} \\
&\overset{(c)}{=} \mathrm{Tr}\left\{\left[\mathbf{F}^{(b)^H}(\mathbf{Q}_{\widehat{\mathbf{H}}}\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H)\mathbf{F}^{(b)}\right]^{-1}\right\} \\
&\overset{(d)}{=} \mathrm{Tr}\left\{\left[\left(\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H\mathbf{F}^{(b)}\right)\left(\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H\mathbf{F}^{(b)}\right)^H\right]^{-1}\right\},
\end{aligned}$$

where (a) results from the invertibility of $\widehat{\mathbf{H}}\mathbf{F}^{(b)}$, (b) follows that $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$ [16], (c) stems from the eigenvalue decomposition (EVD) of $\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}$, and (d) can be obtained by writing the diagonal matrix $\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}$ as $\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}} \times \boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}$. In the EVD $\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} = \mathbf{Q}_{\widehat{\mathbf{H}}}\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H$, $\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}$ is a diagonal square matrix with all eigenvalues of $\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}$ placed along the diagonal in descending order, and $\mathbf{Q}_{\widehat{\mathbf{H}}}$ consists of the corresponding eigenvectors. Note that all the QoS requirements and channel states are contained in this eigenspace since $\widehat{\mathbf{H}} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{H}$, such that $\mathbf{Q}_{\widehat{\mathbf{H}}}$ and $\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}$ are enough for a good RF beamforming design. Although the value of the trace of the inverse of an symmetric matrix, $\mathrm{Tr}[(\cdot)^{-1}]$, can be estimated using a modified Chebyshev algorithm [17], the estimated upper/lower bounds are not tight enough and the algorithm complexity is also too high for the beamforming design application.

Hence, our design is based on an intuition that increasing the amplitudes of the diagonal elements in the matrix $\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H\mathbf{F}^{(b)}$ tends to reduce the value of the trace of the inverse of $\left(\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H\mathbf{F}^{(b)}\right)\left(\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H\mathbf{F}^{(b)}\right)^H$. One validated case is that the inter-user interference is so slight that the matrix $\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H\mathbf{F}^{(b)}$ become diagonal. Note that the k-th entry on the diagonal of $\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H\mathbf{F}^{(b)}$ is the inner product of the k-th column of $\mathbf{Q}_{\widehat{\mathbf{H}}}\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}$ and the k-th column of $\mathbf{F}^{(b)}$. Therefore, we select the codebook bases to match the directions of eigenvectors in steps 5-7 of **Algorithm 1**. To determine a specific column in $\mathbf{F}$, we first pick the corresponding column in $\mathbf{Q}_{\widehat{\mathbf{H}}}\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}$ via mapping the column index in $\mathbf{F}$ to the column index in $\mathbf{F}^{(b)}$. Then the DFT basis vector in codebook $\mathcal{D}_{\mathrm{DFT}}$ with the largest projection onto the picked column in $\mathbf{Q}_{\widehat{\mathbf{H}}}\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}$ will be selected and deleted from the codebook $\mathcal{D}_{\mathrm{DFT}}$. *Herein, we call the basis selection procedures as eigenmode selection*. As we sequentially determine the columns of $\mathbf{F}$, the columns of $\mathbf{Q}_{\widehat{\mathbf{H}}}\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}$ are circularly, from $m = 1, \ldots, K$, visited, since all $\mathbf{F}^{(b)}$'s are designed to match the same $\mathbf{Q}_{\widehat{\mathbf{H}}}\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}$. In such a column-by-column manner, the constraint on $\mathbf{F}$ that M_T is the multiple of K can be dropped, which means there is no limit on the number of RF chains in this design. The complexity of **Algorithm 1** is $\mathcal{O}\left(N_T N_C M_T + K^2 N_T + K^3\right) = \mathcal{O}\left(N_T^2 M_T\right)$, since $N_C = N_T$ and $K \le M_T \le N_T$.

Algorithm 1 Eigenmode Selection Hybrid Beamforming

1: **Initialize** $\boldsymbol{\Lambda}$ in the baseband beamformer using (10)
2: Compute $\widehat{\mathbf{H}} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{H}$ and the EVD $\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} = \mathbf{Q}_{\widehat{\mathbf{H}}}\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}\mathbf{Q}_{\widehat{\mathbf{H}}}^H$
3: **for** $m = 1$: M_T
4: $k = (m-1) \bmod K + 1$
5: pick the k-th column of $\mathbf{Q}_{\widehat{\mathbf{H}}}\boldsymbol{\Sigma}_{\widehat{\mathbf{H}}}^{\frac{1}{2}}$ as $\mathbf{q}_k$
6: **search** all bases $\{\mathbf{d}_n\}$ in $\mathcal{D}_{\mathrm{DFT}}$
7: find a $\mathbf{d}_l$ such that $\mathbf{d}_l = \arg\max_{\mathbf{d}_n} |\mathbf{q}_k^H\mathbf{d}_n|$, where $\mathbf{d}_l \in \mathcal{D}_{\mathrm{DFT}}$
8: **update** $\mathcal{D}_{\mathrm{DFT}} = \mathcal{D}_{\mathrm{DFT}} - \mathbf{d}_l$
9: **end for**
10: Obtain all $\mathbf{d}_l$'s and output $\mathbf{F} = \{\mathbf{d}_l\text{'s}\}$
11: Output $\mathbf{B} = \mathbf{H}_{eff}^H(\mathbf{H}_{eff}\mathbf{H}_{eff}^H)^{-1}\boldsymbol{\Lambda}$, where $\mathbf{H}_{eff} = \mathbf{HF}$.

IV. Simulation Results

In this section, we evaluate the BS power consumption, including the transmission and total power, when the eigenmode selection hybrid beamforming is applied. We next describe our simulation framework, based on MATLAB R2016b, and the simulation results.

A. Simulation Setups

We validate the effectiveness of our proposed approach in the simulations. There are three sets of parameters: i) user channel conditions, ii) QoS (data rate) requirements, and iii) power consumption factors of the BS.

We employ the ULA model for BS antenna array. The large-scale fading factors, β_k's, of all users are uniformly distributed over [0.5, 1.5] to indicate the geometric location difference. The lower and upper bounds of azimuth AoDs $[\phi_{\min}^k, \phi_{\max}^k]$ are set to $[-\pi, \pi]$ for all channels $\tilde{\mathbf{h}}_k$'s. The number of BS antennas N_T is set to 64, while 8-user spatial multiplexing is considered, each of which has a random minimum rate requirement, uniformly distributed over [100 kbps, 30 Mbps]. The bandwidth W is set to 5MHz. The maximum transmission power $P_{tx}^{\max}$ is set to 35W. All other parameters related to the BS power consumption are illustrated in Table I [8], [14].

TABLE I. SIMULATION PARAMETERS FOR BS POWER

Baseband processing power P_{BP}	100W
Circuit power per RF chain $P_{RF,0}$	5W
Maximum BS transmission power $P_{tx}^{\max}$	35W
Peak PA efficiency $\eta_{PA}^{\max}$	$\pi/4$
Feeder efficiency σ_{feed}	50%
Loss factor of DC conversion σ_{DC}	7.5%
Loss factor of main supply σ_{MS}	12%
Loss factor of cooling system σ_{cool}	10%

B. Performance Evaluation

In this paper, the most important performance indicator of the proposed eigenmode selection hybrid beamforming scheme is the BS transmission power given the QoS requirements and channel state information of all users. We also show the corresponding total BS power consumption calculated based on (4). Note that the QoS requirements are assumed to change per second, while channel state information is updated every milli-second so that we evaluate the performance for 1000 times (considering all channel realizations) per QoS set and take the average. The number of RF chains M_T is limited up to 16. Furthermore, since we set a maximum BS transmission power, $P_{tx}^{\max}$, it is possible that the QoS requirements cannot be satisfied under some conditions. For instance, the channel condition is too poor to support the required user rates, called the outage event. Therefore, for a QoS requirement set, we calculate the outage probability as

$$P_{out} = \frac{\text{\# of outage events}}{\text{\# of evaluations per QoS set}} = \frac{\text{\# of outage events}}{1000},$$

which will also be presented in simulation results.

In Fig. 2, the average BS transmission power of eigenmode selection hybrid beamforming is reported given a randomly generated QoS requirement set. The number of dominant propagation paths for each user N_P is set to 4 (solid lines) and 10 (dash lines), representing the poor-scattering and rich-scattering channel conditions respectively. To further confirm the effectiveness of our proposed algorithm, the analog beamsteering scheme in [12] is also applied in the same simulation framework for comparison, which aims to pursue the physical propagation paths with RF beamforming instead of eigenmodes while ZF is also used for baseband beamforming. From Fig. 2 we observe that our proposed eigenmode selection scheme needs significantly less transmission power to meet the QoS requirements compared to the analog beamsteering scheme. The performance gap between these two schemes (for same number of paths) becomes narrow when the number of RF chains increases from 8 to 16, as the analog beamsteering scheme benefits from the fact that there are more columns in RF beamformer to catch the channel gains of physical propagation paths. However, for the eigenmode selection scheme, most of the channel gains can be harvested in the eigenspace even with a fewer number of RF chains, such as 8, so that increase on RF chains has less evident effect. Nevertheless, when it transits from the 4-path case to the 10-path case, the eigenmode selection scheme outperforms the analog beamsteering prominently, since the analog beamsteering scheme can only harvest the channel gains from a limited portion of all propagation paths, while the eigenmode selection scheme can catch a higher channel gain by projecting multiple physical paths onto few dominant eigenvectors to be captured by a limited number of RF chains. Briefly, the eigenmode selection scheme can work effectively in all types of channels, especially in the rich-scattering environments.

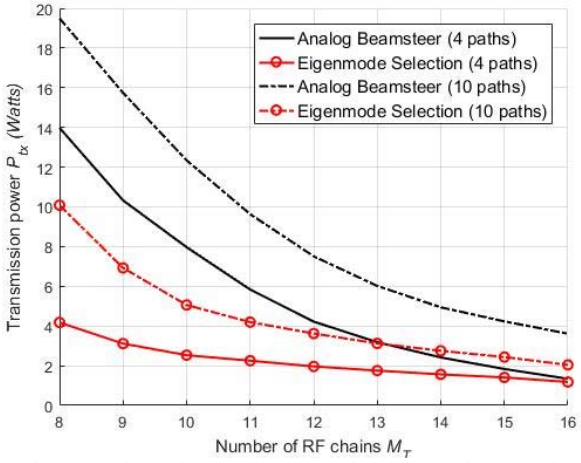

Fig. 2. The comparison of average BS transmission power between eigenmode selection and analog beamsteering [12] with an increasing number of RF chains when the number of propagation paths in each user channel is set to 4 and 10

We also present the average total BS power consumption comparison with the same simulation setups (only for 10-path case and the number of RF chains M_T is chosen as 8, 10, 12, 14 and 16) in Table II. Significant average power savings can be obtained, though the power savings decreases with increasing number of RF chains. Even the average total BS power begins to increase when M_T is large, e.g., 16. This is due to the increasing circuit power of more RF chains. Moreover, the percentage of BS transmission power that the eigenmode selection scheme can saved, compared with the analog steering scheme, is also getting lower, which means the circuit power of RF chains gradually dominates the BS power consumption when M_T is large. Therefore, the optimal number of RF chains for the sake of saving BS power is a value located between its minimum and maximum, which leads to a very interesting problem of how to balance the increase of the circuitry power of RF chains and the decrease of the transmission power through setting an optimal number of RF chains.

TABLE II. AVERAGE TOTAL BS POWER COMPARISON BETWEEN EIGENMODE SELECTION AND ANALOG BEAMSTEERING UNDER DIFFERENT NUMBERS OF RF CHAINS

Number of RF chains	8	10	12	14	16
Analog beamsteering	457.22W	373.23W	320.91W	299.48W	295.02W
Eigenmode selection	328.75W	273.77W	267.85W	269.61W	273.54W
BS power saving	28.1%	26.65%	16.53%	9.97%	7.28%

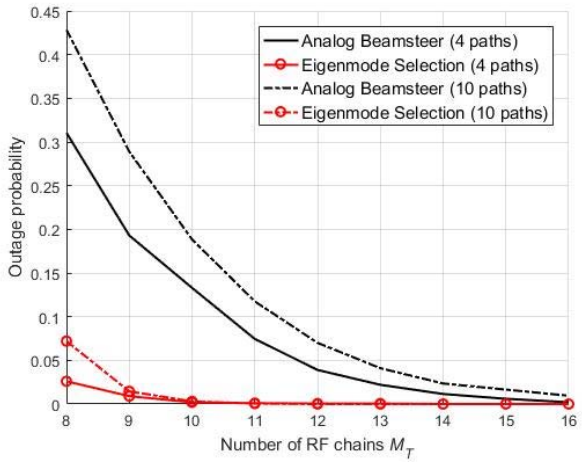

Fig. 3. The comparison of outage probability between eigenmode selection and analog beamsteering [12] with an increasing number of RF chains when the number of propagation paths in each user channel is set to 4 and 10.

Furthermore, we show another advantage of our eigenmode selection scheme with respect to the outage probability performance in Fig. 3, which reveals another critical problem other than saving BS power: the effective lower bound of the number of RF chains. As shown in Fig. 3, the analog beamsteering scheme is unable to meet the QoS requirements with an unacceptably high probability given that the maximum BS transmission power $P_{tx}^{max} = 35\text{W}$, while the eigenmode selection scheme can consistently achieve a very low outage probability. With the analog beamsteering scheme in [12], effective communications (e.g., $P_{out} < 0.05$) only happens when the number of RF chains is larger than 13. However, for the eigenmode selection scheme, the effective lower bound of the number of RF chains could be down to 8 or 9, which demonstrates that the eigenmode selection has more valid settings with fewer RF chains in terms of saving BS power.

V. Conclusion

In this paper, the fully-connected hybrid beamforming architecture is investigated for minimizing the BS transmission power when the user QoS requirements are satisfied in a massive MU-MIMO downlink communication scenario. An eigenmode selection hybrid beamforming scheme is proposed to choose the optimal bases in DFT codebook to pursue the channel gains in the corresponding eigenspace. We also present a phase array batch-switching structure to realize the eigenmode selection beamforming economically. Using simulation results, we demonstrate that the proposed eigenmode selection scheme can achieve significant power savings while satisfying users' data rate requirements, especially in the rich-scattering environment, even with few number of RF chains. As a future extension of this work, we plan to investigate how we can adaptively switch RF chains on/off such that the long-term average power consumption can be minimized for the BS with eigenmode selection hybrid beamforming.

VI. Acknowledgement

This work was supported in part by the Center for Wireless Communications at UC San Diego.

References

[1] J. G. Andrews *et al.*, "What will 5G be," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Oct. 2014.

[3] A. F. Molisch *et al.*, "Hybrid Beamforming for Massive MIMO – A Survey," in *arXiv: 1609.05078*, Sept. 2016.

[4] O. E. Ayach, R. W. Heath, S. Rajagopal and Z. Pi, "Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays," in *IEEE Global Commun. Conf.*, pp. 3676–3480, Dec. 2013.

[5] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[6] W. Ni, X. Dong and W.-S. Lu, "Near-optimal hybrid processing for massive mimo systems via matrix decomposition," *arXiv:1504.03777*, Apr. 2015.

[7] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, pp. 201–211, Jan. 2016.

[8] R. Guruprasad, K. Son and S. Dey, "Power-efficient base station operation through user QoS-aware adaptive RF chain switching technique," in *Proc. of IEEE Int. Commun Conf. (ICC)*, pp. 244–250, Jun. 2015.

[9] H. Yang and T. L. Marzetta, "Energy efficient design of massive MIMO: How many antennas ?," in *IEEE Vehicular Tech. Conf. (VTC Spring)*, pp. 1–5, 2015.

[10] W. Liu, S. Han and C. Yang, "Energy Efficiency Scaling Law of Massive MIMO Systems," *IEEE Trans Commun.*, vol. 65, no. 1, pp. 107–121, Sept. 2016.

[11] E. Bjornson, L. Sanguinetti, J. Hoydis and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?," *IEEE Trans Wireless Commun.*, vol. 14, pp. 3059-–3075, June 2015.

[12] C. Lin and G. Y. Li, "Energy-efficient design of indoor mmWave and sub-THz systems with antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 4660–4672, July 2016.

[13] R. M.-Rial, C. Rusu, N. G.-Prelcic, A. Alkhateeb and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase Shifters or Switches?" *IEEE Access*, vol. 4, pp. 247-267, Jan. 2016.

[14] H. Holtkamp, G. Auer, V. Giannini and H. Haas, "A parameterized base station power model," vol. 17, no. 11, Nov. 2013.

[15] A. H. Mehana and A. Nosratinia, "Diversity of MIMO linear precoding," *IEEE Trans. Info. Theory*, vol. 60, no. 2, Feb. 2014.

[16] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Technical University of Denmark, Nov. 2012.

[17] G. Meurant, "Estimates of the trace of the inverse of a symmetric matrix using the modified Chebyshev algorithm," in *Numerical Algorithms*, vol. 51, no. 3, pp. 309–318, July 2009.