Hybrid Multi-User Precoding with Amplitude and Phase Control

Miguel R. Castellanos†, Vasanthan Raghavan⋆, Jung H. Ryu⋆, Ozge H. Koymen⋆, Junyi Li⋆,
David J. Love†, and Borja Peleato†
†Purdue University, West Lafayette, IN 47907, USA
⋆Qualcomm Corporate R&D, Bridgewater, NJ 08807, USA

***Abstract*— There has been a strong interest in understanding hybrid precoding tradeoffs for millimeter wave (mmW) multi-input multi-output (MIMO) systems. A common assumption in most of these works is that the analog part of the hybrid precoder can only be designed with phase shifters. The consequent search for analog and digital precoding matrices is solved with different *black box*-type optimization algorithms. In contrast, this work motivates an analog precoding structure at the base-station end that can be realized with *both* phase shifters and gain controls. Such a structure is necessary for interference management in multi-user transmissions and is easily realized with low complexity and cost. We then propose a feedback framework of the top-P beams over a beam alignment phase from each user. This framework allows the base-station to reconstruct the channel matrix between it and each user, and to manage interference with a simple zeroforcing solution. Such a structured solution is implemented with the amplitude and phase control of the analog part of the hybrid precoder. We finally illustrate the performance improvement with the proposed solution over simpler constructions such as beam steering that can be implemented with phase shifters alone.**

***Index Terms*— Millimeter wave systems, beamforming, hybrid precoding, phase shifter, amplitude control, multi-user systems**

I. Introduction

With the increased interest in the commercialization of millimeter wave (mmW) systems, a number of studies have appeared on the channel quality at mmW carrier frequencies relative to sub-6 GHz systems [1]–[3]. These studies show that while mmW systems suffer from marginally increased path losses and substantially increased penetration and blockage losses, by focussing on small cell coverage (100-500 m coverage), using large antenna arrays at the base-station end (64-256 antennas) and dual-polarized subarray diversity (4-8 subarrays), significant rate improvements can be realized in practice at a reasonable cost.

To obtain this spatial array gain, mmW systems leverage spatial *sparsity* corresponding to the few dominant clusters in the channel [1], [2], [4] by focussing on *directional* beamforming solutions [5]–[12]. Spatial sparsity of the channel along with the use of large antenna arrays motivates a subset of physical layer beamforming schemes based on *directional* transmissions for signaling. In this context, there have been a number of studies on the design and performance analysis of directional beamforming/precoding structures for single-user multi-input multi-output (MIMO) systems. These works [8]–[11] show that directional schemes are not only good from an implementation standpoint, but are also robust to phase changes across clusters and allow a smooth tradeoff between peak beamforming gain and initial user discovery latency. There has also been progress in generalizing such directional constructions for multi-user MIMO transmissions [12]–[14].

While legacy systems use as many radio frequency (RF) chains[1] as the number of antennas, their higher cost, energy consumption, area and weight at millimeter wave carrier frequencies has resulted in the popularity of *hybrid beamforming* systems [15]–[17]. Spatial sparsity of millimeter wave channels ensures that having as many RF chains as the number of dominant clusters in the channel is sufficient to reap the full array gain possible over these channels. A number of recent works have addressed hybrid beamforming for millimeter wave systems. The problem of finding the optimal precoder and combiner with a hybrid architecture is posed as a sparse reconstruction problem in [9], leading to algorithms and solutions based on basis pursuit methods. While the solutions achieve good performance in certain cases, to address the performance gap between the solution proposed in [9] and the unconstrained beamformer structure, an iterative scheme is proposed in [18], [19] relying on a hierarchical training codebook for adaptive estimation of millimeter wave channels. In [20], it is established that a hybrid architecture can approach the performance of a digital architecture as long as the number of RF chains is twice that of the data-streams. A heuristic algorithm with good performance is developed when this condition is not met. Other works such as [21]–[23] have also explored iterative/algorithmic solutions for hybrid beamforming.

A common theme that underlies most of these works is the assumption of *phase-only* control in the RF/analog domain of the hybrid beamforming architecture. This assumption makes sense at the user end with a small number of antennas (relative to the base-station end), where operating the PAs below their peak rating across RF chains can lead to a substantially poor uplink performance. On the other hand, amplitude control (denoted as *amplitude tapering* in the antenna theory literature) is necessary at the base-station end with a large number of antennas for side-lobe management and mitigating out-of-band emissions. Further, given that the base-station is a network resource, simultaneous amplitude and phase control of the individual antennas across RF chains is feasible at millimeter wave base-stations at a low-complexity and cost [24, pp. 285-289], [25], [26]. In particular, the millimeter wave experimental prototype demonstrated in [27] allows simultaneous amplitude and phase control enabled by advanced calibration techniques.

[1]An RF chain includes (but is not limited to) analog-to-digital converters, digital-to-analog converters, mixers, low-noise and power amplifiers (PAs), etc.

Thus, it is important to consider a hybrid architecture with these constraints. Further, given the directional structure in the channel, a *black box*-type iterative or algorithmic solution that does not provide an intuitive description of the beam weights is less preferable over a solution that is constructed out of measurement reports obtained over an initial *beam alignment* phase with a directional structure for the sounding beams.

Main Contributions: With this backdrop, this work addresses these two fundamental issues of amplitude and phase control, and a constructive feedback scheme design for hybrid beamforming. It is assumed that the base-station trains all the users in the cell with a cell-specific codebook of beamforming vectors over an initial beam alignment phase. Each user makes an estimate of the top-P (where $P \geq 1$) beams over this phase and reports the beam indices to be used by the base-station as well as the measured/received signal-to-noise ratios (SNRs). The simplest implementation at the base-station uses only the best beam information for beam steering or zeroforcing as in [13], [14], with other beams serving as fall back options. In contrast to this approach, we propose to *reconstruct* or *estimate* a rank-P approximation of the channel matrix between the base-station and the user (at the base-station end). To realize this scheme, we propose that the users communicate the phase estimates of the top-P beams back to the base-station, as well as the cross-correlation information of the top-P beams at the user end with the beam subsequently used in multi-user reception. Leveraging the rank-P channel approximation, we propose the use of a zeroforcing structure that is then quantized to meet the RF precoding constraints (amplitude and phase control) at the base-station end for simultaneous transmissions.

To benchmark and compare the performance of the proposed scheme, we establish an upper bound for the sum rate. This is a fundamentally difficult problem given the non-convex dependence of the sum rate on the beamforming vectors [28]. This bound is based on an intuitive understanding of the zeroforcing structure. Numerical studies show that the proposed scheme performs significantly better than a naïve beam steering solution even for an initial beam alignment codebook of poor resolution. Further, the proposed scheme is comparable with the established upper bound provided the beam alignment codebook resolution is moderate-to-good. Thus, our work establishes the utility and efficacy of the proposed feedback techniques.

Note: Proofs of all the statements are found in [29].

II. System Setup

We consider a narrowband millimeter wave system in a downlink scenario with a single base-station serving K users. We assume that the base-station and each user are equipped with uniform planar arrays containing $N_t = N_{\mathsf{tx}} \times N_{\mathsf{tz}}$ antennas and $N_r = N_{\mathsf{rx}} \times N_{\mathsf{rz}}$ antennas, respectively, with half-wavelength element spacing at both ends. We focus on the case where the base-station is equipped with $M_t \leq N_t$ RF chains and the user is equipped with only one RF chain. In a practical use-case of interest for mmW systems where each user has two layers/RF chains, but uses these two layers for polarization-based transmissions over a single spatial layer, the setup considered in this paper is relevant. In the case of more than a single spatial layer at the users, the proposed framework can be easily extended.

Denoting the $N_r \times N_t$ channel matrix between the base-station and the k-th user as $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$, the extended Saleh-Valenzuela geometric model leads to the following setup,

$$\mathbf{H}_k = \sqrt{\frac{N_r N_t}{L_k}} \sum_{\ell=1}^{L_k} \alpha_{k,\ell}\, \mathbf{u}_{k,\ell}\, \mathbf{v}_{k,\ell}^{\dagger}, \tag{1}$$

where L_k denotes the number of clusters/paths, $\alpha_{k,\ell}$ denotes the cluster gain, and $\mathbf{u}_{k,\ell}$ and $\mathbf{v}_{k,\ell}$ denote the receive and transmit array steering vectors, respectively. The normalization in (1) ensures that $\mathcal{E}\left[\mathsf{Tr}(\mathbf{H}_k \mathbf{H}_k^{\dagger})\right] = N_r N_t$.

We assume that the base-station schedules as many users as it has RF chains and precodes a single spatial layer to each scheduled user. In particular, the base-station precodes the data symbol s_m to the m-th user with a digital $K \times 1$ beamformer $\mathbf{f}_{\mathsf{D},\,m}$. Upconversion of the baseband data to the carrier frequency is then performed by an $N_t \times K$ RF precoder $\mathbf{F}_{\mathsf{RF}}$. The system model for the k-th user is thus given as

$$\mathbf{y}_k = \sqrt{\frac{\rho}{K}} \mathbf{H}_k \mathbf{F}_{\mathsf{RF}} \cdot \left[\sum_{m=1}^{K} \mathbf{f}_{\mathsf{D},\,m} s_m \right] + \mathbf{n}_k, \tag{2}$$

where ρ is the pre-precoding SNR and $\mathbf{n}_k$ is the $N_r \times 1$ additive white Gaussian noise vector satisfying $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$. For simplicity, we assume that $s_k \sim \mathcal{CN}(0,1)$. A simple realization of the hybrid precoding architecture can be obtained by letting $[\mathbf{f}_{\mathsf{D},1}, \ldots, \mathbf{f}_{\mathsf{D},K}] = \mathbf{I}_K$ (or use them for wideband adaptation, which is not considered in this work) and denoting the k-th column of the RF precoder $\mathbf{F}_{\mathsf{RF}}$ by $\mathbf{f}_k$. We will make this simplifying assumption through the rest of this work. In this case, the transmit model in (2) can be simplified as

$$\mathbf{y}_k = \sqrt{\frac{\rho}{K}}\, \mathbf{H}_k \left[\mathbf{f}_1 \cdots \mathbf{f}_K\right] \cdot \left[s_1 \cdots s_K\right]^T + \mathbf{n}_k\,. \tag{3}$$

Each user processes incoming signals with an $N_r \times 1$ user-specific combining vector to create an estimate of the transmitted symbol

$$\widehat{s_k} = \mathbf{g}_k^{\dagger}\, \mathbf{y}_k = \sqrt{\frac{\rho}{K}} \mathbf{g}_k^{\dagger} \mathbf{H}_k \mathbf{f}_k s_k + \sqrt{\frac{\rho}{K}} \sum_{m \neq k} \mathbf{g}_k^{\dagger} \mathbf{H}_k \mathbf{f}_m s_m + \mathbf{g}_k^{\dagger}\, \mathbf{n}_k\,. \tag{4}$$

The achievable rate at the k-th user by treating multi-user interference as noise is given as

$$\mathcal{R}_k = \log\left(1 + \mathsf{SINR}_k\right), \tag{5}$$

where SINR_k denotes the signal-to-interference and noise ratio of the k-th user, defined as,

$$\mathsf{SINR}_k \triangleq \frac{\frac{\rho}{K} \cdot |\mathbf{g}_k^{\dagger} \mathbf{H}_k \mathbf{f}_k|^2}{1 + \frac{\rho}{K} \cdot \sum_{m \neq k} |\mathbf{g}_k^{\dagger} \mathbf{H}_k \mathbf{f}_m|^2}. \tag{6}$$

Motivated by practical beamforming architectures, the base-station is assumed to have finite-precision control over both the phases and amplitudes of the entries in $\mathbf{f}_k$. Capturing this assumption, the phase and amplitude values of $\mathbf{f}_k$ are taken from a $2^{B_{\mathsf{phase}}}$ and $2^{B_{\mathsf{amp}}}$ set of quantized values, as below:

$$\angle \mathbf{f}_k(i) \in \left\{\phi_1, \ldots, \phi_{2^{B_{\mathsf{phase}}}}\right\}, \quad |\mathbf{f}_k(i)| \in \left\{A_1, \ldots, A_{2^{B_{\mathsf{amp}}}}\right\}, \tag{7}$$

where $0 \leq A_1 < A_2 < \cdots < A_{2^{B_{\mathsf{amp}}}}$. With these constraints on the precoder structure, the transmit power constraint is given by $\sum_{k=1}^{K} \mathbf{f}_k^\dagger \mathbf{f}_k \leq K$. Note that prior works on hybrid precoding only consider the use of phase shifter control for $\mathbf{F}_{\mathsf{RF}}$. In the following section, we develop a scheme to design multi-user beams $\mathbf{f}_k$ to improve the sum rate $\mathcal{R}_{\mathsf{sum}} \triangleq \sum_{k=1}^{K} \mathcal{R}_k$.

III. Multi-User Beamformer Design

The focus of this section is to develop an advanced feedback mechanism and a systematic design of the multi-user beamforming structure based on a directional representation of the channel. This structure allows the base-station to combat multi-user interference in simultaneous transmissions.

In practical mmW systems, an initial acquisition process, commonly referred to as the *beam alignment phase*, is performed to identify a coarse beam at both the base-station and user ends. This is followed by a *beam refinement phase* where the identified beams are fine-tuned in order to lead to a high-rate link. This entire process can be modeled as a beam sweep performed by the base-station and the users over codebooks, $\mathcal{F}_{\mathsf{tr}}$ and $\mathcal{G}_{\mathsf{tr}}^k$, respectively. Let the base-station be equipped with an N element beam training codebook $\mathcal{F}_{\mathsf{tr}} = \{\mathbf{f}_{\mathsf{tr},1}, \cdots, \mathbf{f}_{\mathsf{tr},N}\}$ which is known only to the base-station. Correspondingly, the k-th user is equipped with an M element training codebook $\mathcal{G}_{\mathsf{tr}}^k = \left\{\mathbf{g}_{\mathsf{tr},1}^{(k)}, \cdots, \mathbf{g}_{\mathsf{tr},M}^{(k)}\right\}$ which is known only to the k-th user.

During the beam training phase, the base-station and each user run through their codebooks to approximate the received SNR, defined as,

$$\mathsf{SNR}_{\mathsf{rx}}^{(k)}(m,n) \triangleq \left| \left(\mathbf{g}_{\mathsf{tr},m}^{(k)}\right)^\dagger \mathbf{H}_k \mathbf{f}_{\mathsf{tr},n} \right|^2. \tag{8}$$

The top-P beam index pairs for the k-th user that maximize the received SNR (in non-increasing order) are denoted as

$$\mathcal{M} = \left\{(m_1^k, n_1^k), \cdots, (m_P^k, n_P^k)\right\} \tag{9}$$

and are fed back. Further, we denote the realized received SNR with $\mathbf{f}_{\mathsf{tr},n_\ell^k}$ and $\mathbf{g}_{\mathsf{tr},m_\ell^k}^{(k)}$ as $\mathsf{SNR}_{\mathsf{rx},\ell}^{(k)}$.

A typical use of the top/best beam feedback in multi-user transmissions is to construct advanced multi-user beams which serve a certain objective. This information could be simplistically used to combat blockage or fading with fall back options. Alternately, in [14], the proposed multi-user schemes leverage the top beam feedback to combat interference induced by simultaneously scheduled users via a zeroforcing or a generalized eigenvector solution. In this work, we develop a scheme which leverages the top-P beam pair indices acquired in the beam training phase. Due to the spatial sparsity of mmW channels, directional information obtained from the beam training phase is sufficient to generate a rank-P approximation of $\mathbf{H}_k$. The base-station aims to acquire an estimate of $\mathbf{H}_k$ through feedback as

$$\widehat{\mathbf{H}}_k = \sum_{\ell=1}^{P} \widehat{\alpha}_{k,\ell}\, \widehat{\mathbf{u}}_{k,\ell}\, \widehat{\mathbf{v}}_{k,\ell}^\dagger. \tag{10}$$

In (10), $\widehat{\mathbf{u}}_{k,\ell}$ and $\widehat{\mathbf{v}}_{k,\ell}$ are estimates of the array steering vectors $\mathbf{u}_{k,\ell}$ and $\mathbf{v}_{k,\ell}$, respectively, and $\widehat{\alpha}_{k,\ell}$ is an estimate of $\alpha_{k,\ell}$.

Given the channel model structure in (1), (10) is simplified by estimating $\mathbf{v}_{k,\,\ell}$ and $|\alpha_{k,\,\ell}|$ by $\mathbf{f}_{\mathsf{tr},n_\ell^k}$ and $\gamma_{k,\ell}$, respectively, where

$$\gamma_{k,\ell} \triangleq \sqrt{\mathcal{Q}_{B_{\mathsf{SNR}}}\left(\mathsf{SNR}_{\mathsf{rx},\,\ell}^{(k)}\right)} \tag{11}$$

for some choice of B_{SNR}. In (11), $\mathcal{Q}_B(\cdot)$ denotes an appropriately-defined B-bit quantization operation of the quantity under consideration. However, estimating $\widehat{\mathbf{H}}_k$ as in (10) is not complete until we have an estimate for $\angle\alpha_{k,\ell}$ and $\mathbf{u}_{k,\ell}$. For $\angle\alpha_{k,\ell}$, we define $\varphi_{k,\ell}$ as the $B_{\mathsf{est,\,phase}}$-bit quantization of the phase of an estimate $\widehat{\mathbf{s}}_{\mathsf{tr},k,\ell}$ of a pilot symbol $\mathbf{s}_{\mathsf{tr},k,\ell}$

$$\varphi_{k,\ell} \triangleq \mathcal{Q}_{B_{\mathsf{est,\,phase}}}\left(\angle\widehat{\mathbf{s}}_{\mathsf{tr},k,\ell}\right), \tag{12}$$

where

$$\widehat{\mathbf{s}}_{\mathsf{tr},k,\ell} = \left(\mathbf{g}_{\mathsf{tr},\,m_\ell^k}^{(k)}\right)^\dagger \left[\sqrt{\rho}\, \mathbf{H}_k \mathbf{f}_{\mathsf{tr},\,n_\ell^k} \mathbf{s}_{\mathsf{tr},k,\ell} + \mathbf{n}_{k,\ell}\right] \tag{13}$$

for some choice of $B_{\mathsf{est,\,phase}}$. The noise term $\mathbf{n}_{k,\ell}$ captures the additive noise in the initial beam alignment process corresponding to the ℓ-th best beam pair.

For $\mathbf{u}_{k,\ell}$, we note that the base-station not only needs the beam indices $\{m_\ell^k\}$ that are useful for the user side, but also the useful part of the user's codebook ($\mathcal{G}_{\mathsf{tr}}^k$) since the base-station is typically not aware of it. To avoid this unnecessary complexity and feedback given the proprietary nature of $\mathcal{G}_{\mathsf{tr}}^k$, we assume that the k-th user uses a multi-user reception beam $\mathbf{g}_k$. In the simplest manifestation, $\mathbf{g}_k$ could be the best training beam learned in the beam alignment phase, $\mathbf{g}_{\mathsf{tr},m_1^k}^{(k)}$. However, a more sophisticated choice for $\mathbf{g}_k$ is not precluded. We then note that the estimated SINR, defined as,

$$\widehat{\mathsf{SINR}}_k \triangleq \frac{\frac{\rho}{K} \cdot |\mathbf{g}_k^\dagger \widehat{\mathbf{H}}_k \mathbf{f}_k|^2}{1 + \frac{\rho}{K} \cdot \sum_{m \neq k} |\mathbf{g}_k^\dagger \widehat{\mathbf{H}}_k \mathbf{f}_m|^2} \tag{14}$$

is only dependent on $\widehat{\mathbf{H}}_k$ in the form of $\mathbf{g}_k^\dagger \widehat{\mathbf{H}}_k$. Building on this fact, each user generates $\{\beta_{k,\,\ell}\}$, defined as,

$$\beta_{k,\,\ell} \triangleq \mathbf{g}_k^\dagger \widehat{\mathbf{u}}_{k,\ell} \ \text{ where } \ \widehat{\mathbf{u}}_{k,\ell} = \mathbf{g}_{\mathsf{tr},\,m_\ell^k}^{(k)}. \tag{15}$$

It then quantizes the amplitude and phase of $\beta_{k,\ell}$ for some choice of $B_{\mathsf{corr,\,amp}}$ and $B_{\mathsf{corr,\,phase}}$ and feeds them back

$$\mu_{k,\ell} \triangleq \mathcal{Q}_{B_{\mathsf{corr,\,amp}}}\left(|\beta_{k,\ell}|\right), \quad \nu_{k,\ell} \triangleq \mathcal{Q}_{B_{\mathsf{corr,\,phase}}}\left(\angle\beta_{k,\ell}\right). \tag{16}$$

For both $\varphi_{k,\ell}$ and $\nu_{k,\ell}$, without loss in generality, relative phases with respect to $\varphi_{k,1}$ and $\nu_{k,1}$ (that is, $\varphi_{k,\ell} - \varphi_{k,1}$ and $\nu_{k,\ell} - \nu_{k,1}$) can be reported. Following the above discussion, the base-station approximates $\mathbf{g}_k^\dagger \widehat{\mathbf{H}}_k$ as

$$\mathbf{g}_k^\dagger \widehat{\mathbf{H}}_k = \sum_{\ell=1}^{P} \mu_{k,\ell}\, \gamma_{k,\ell} \cdot e^{j(\varphi_{k,\ell} + \nu_{k,\ell})} \cdot \left(\mathbf{f}_{\mathsf{tr},n_\ell^k}\right)^\dagger. \tag{17}$$

From (17), the phases appear in the channel reconstruction in the form $\varphi_{k,\ell} + \nu_{k,\ell}$ and thus their feedback overhead can be combined (denoted as B_{phase}). Thus, the net feedback overhead for the proposed scheme is given as

$$\begin{aligned} B_{\mathsf{feedback}} = P \cdot &\left[\log_2(N) + B_{\mathsf{SNR}} + B_{\mathsf{corr,\,amp}}\right] \\ &+ (P-1) \cdot B_{\mathsf{phase}} \ \ \text{(in bits)}. \end{aligned} \tag{18}$$

The base-station uses the channel matrix constructed for each user based on its feedback information ($\mathbf{g}_k^\dagger \widehat{\mathbf{H}}_k$) and generates a good beamformer structure, illustrated in the next result, for use in multi-user transmissions.

Proposition 1. *The zeroforcing beamformer structure is one where for every user that is simultaneously scheduled, the beam $\mathbf{f}_k$ nulls the multi-user interference in $\widehat{\mathsf{SINR}}_m$, $m \neq k$ with $\widehat{\mathsf{SINR}}_m$ as given in (14). The beams $\{\mathbf{f}_m\}$ in the zeroforcing structure are the unit-norm column vectors of the $N_{\mathsf{t}} \times K$ matrix $\mathcal{H}^\dagger \left(\mathcal{H}\mathcal{H}^\dagger\right)^{-1}$, where $\mathcal{H}$ is the $K \times N_{\mathsf{t}}$ matrix given as*

$$\mathcal{H} = \begin{bmatrix} \mathbf{g}_1^\dagger \widehat{\mathbf{H}}_1 \\ \mathbf{g}_2^\dagger \widehat{\mathbf{H}}_2 \\ \vdots \\ \mathbf{g}_K^\dagger \widehat{\mathbf{H}}_K \end{bmatrix}. \tag{19}$$

□

We are interested in benchmarking the performance of the zeroforcing structure against an upper bound on $\mathcal{R}_{\mathsf{sum}}$. This problem is a non-convex optimization [28] that appears to be complicated. In this context, an alternate formulation based on the signal-to-leakage and noise ratio metric that *simultaneously* maximizes the array gain seen by the k-th user, $|\mathbf{g}_k^\dagger \mathbf{H}_k \mathbf{f}_k|^2$, and minimizes the interfering array gain seen by the other users, $|\mathbf{g}_m^\dagger \mathbf{H}_m \mathbf{f}_k|^2$, $m \neq k$ is relevant. Since these objectives are in some sense conflicting and can be weighed differently, we consider the composite metric

$$\mathsf{SLNR}_k \triangleq \frac{\eta_{k,k}\, |\mathbf{g}_k^\dagger \mathbf{H}_k \mathbf{f}_k|^2}{1 + \sum_{m \neq k} \eta_{m,k}\, |\mathbf{g}_m^\dagger \mathbf{H}_m \mathbf{f}_k|^2} \tag{20}$$

for an appropriate set of weighting factors $\eta_{m,k} \geq 0$ with $m, k \in \{1, \cdots, K\}$.

Building on Prop. 1, we now develop an upper bound for $\mathcal{R}_{\mathsf{sum}}$ motivated by the zeroforcing structure. In this direction, we consider a signal-to-leakage-type metric equivalent to (20) based on the estimated channel matrix $\widehat{\mathbf{H}}_k$

$$\widehat{\mathsf{SLNR}}_k \triangleq \frac{\eta_{k,k}\, |\mathbf{g}_k^\dagger \widehat{\mathbf{H}}_k \mathbf{f}_k|^2}{1 + \sum_{m \neq k} \eta_{m,k}\, |\mathbf{g}_m^\dagger \widehat{\mathbf{H}}_m \mathbf{f}_k|^2} \tag{21}$$

for an appropriate set of weighting factors $\eta_{m,k} \geq 0$ with $m, k \in \{1, \cdots, K\}$.

Proposition 2. *Assuming that $\{\widehat{\mathbf{H}}_m^\dagger \mathbf{g}_m\}$ and $\{\eta_{m,k}\}$ are known at the base-station, the choice of $\mathbf{f}_k$ that maximizes $\widehat{\mathsf{SLNR}}_k$ is given by the generalized eigenvector structure*

$$\mathbf{f}_k = \frac{\left(\mathbf{I}_{N_{\mathsf{t}}} + \sum_{m \neq k} \eta_{m,k}\, \widehat{\mathbf{H}}_m^\dagger \mathbf{g}_m \mathbf{g}_m^\dagger \widehat{\mathbf{H}}_m\right)^{-1} \widehat{\mathbf{H}}_k^\dagger \mathbf{g}_k}{\left\| \left(\mathbf{I}_{N_{\mathsf{t}}} + \sum_{m \neq k} \eta_{m,k}\, \widehat{\mathbf{H}}_m^\dagger \mathbf{g}_m \mathbf{g}_m^\dagger \widehat{\mathbf{H}}_m\right)^{-1} \widehat{\mathbf{H}}_k^\dagger \mathbf{g}_k \right\|}. \tag{22}$$

□

While it is hard to simplify $\mathbf{f}_k$ in (22), it can be seen that $\mathbf{f}_k$ can be expressed as

$$\mathbf{f}_k = \frac{\sum_{m=1}^K \widehat{\delta}_{m,k} \widehat{\mathbf{H}}_m^\dagger \mathbf{g}_m}{\left\| \sum_{m=1}^K \widehat{\delta}_{m,k} \widehat{\mathbf{H}}_m^\dagger \mathbf{g}_m \right\|} \tag{23}$$

for some complex scalars $\widehat{\delta}_{m,k}$. In other words, the optimal $\mathbf{f}_k$ is in the span of $\{\widehat{\mathbf{H}}_m^\dagger \mathbf{g}_m\}$ with the weights $\{\widehat{\delta}_{m,k}\}$ that make the linear combination being a complicated function of $\{\eta_{m,k}\}$ as well as $\{\widehat{\mathbf{H}}_m^\dagger \mathbf{g}_m\}$. With this interpretation, while Prop. 2 considers only the maximization of $\widehat{\mathsf{SLNR}}_k$ (not even the sum rate with $\widehat{\mathbf{H}}_k$), we can consider the optimization of $\mathcal{R}_{\mathsf{sum}}$ over $\mathbf{f}_k$ from a class $\mathcal{F}_k$, defined as,

$$\mathcal{F}_k \triangleq \Bigg\{ \mathbf{f}_k \;:\; \mathbf{f}_k = \frac{\sum_{n=1}^N \delta_{n,k} \mathbf{f}_{\mathsf{tr},\, n}}{\left\| \sum_{n=1}^N \delta_{n,k} \mathbf{f}_{\mathsf{tr},\, n} \right\|} \\ \text{such that } \delta_{n,k} \in \mathbb{C},\; k = 1, \cdots, K \Bigg\}. \tag{24}$$

Theorem 1. *Assume that the same multi-user beams $\mathbf{g}_k$ as in the zeroforcing scheme are used for reception at the k-th user. Let $\{\delta_{n,k}^\star\}$ be defined as the solution to the search over the complex scalars $\{\delta_{n,k}\}$*

$$\{\delta_{n,k}^\star\} = \arg \max_{\{\delta_{n,k} \,:\, \mathbf{f}_k \,\in\, \mathcal{F}_k\}} \mathcal{R}_{\mathsf{sum}}. \tag{25}$$

With $\mathbf{g}_k$ as above and

$$\mathbf{f}_k = \frac{\sum_{n=1}^N \delta_{n,k}^\star \mathbf{f}_{\mathsf{tr},\, n}}{\left\| \sum_{n=1}^N \delta_{n,k}^\star \mathbf{f}_{\mathsf{tr},\, n} \right\|}, \tag{26}$$

we obtain an upper bound to the sum rate with the zeroforcing scheme. This upper bound to the sum rate is computed using (5) and (6) with $\mathbf{f}_k$ and $\mathbf{g}_k$ as above. □

IV. Numerical Studies

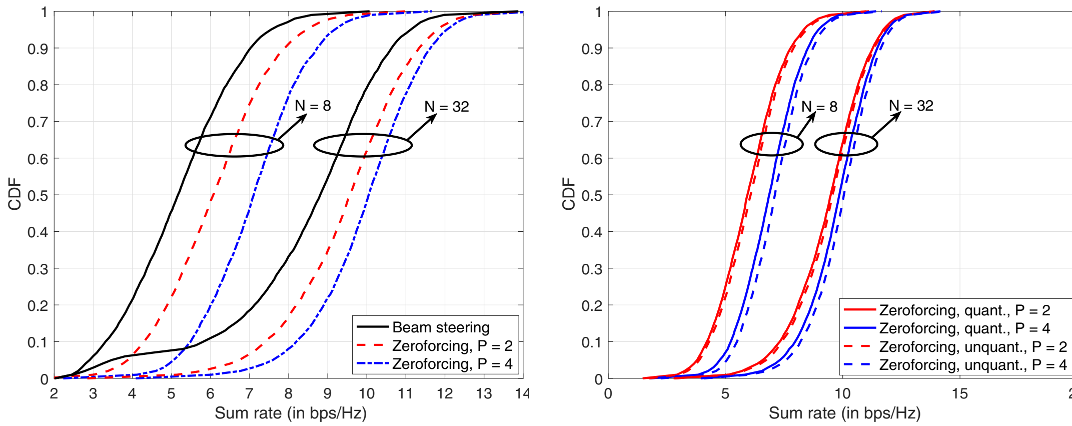
In this section, we present numerical studies in a single-cell downlink framework to illustrate the merits of the proposed beamforming scheme and compare it with the *unrealizable* upper bound of Theorem 1. In our studies, the antenna array dimensions are assumed to be $N_{\mathsf{tx}} = 16$ and $N_{\mathsf{tz}} = 4$ at the base-station end, and $N_{\mathsf{rx}} = 2$ and $N_{\mathsf{rz}} = 2$ at each user. The channel model from (1) is used to generate a channel matrix with $L_k = 6$ clusters, AoDs uniformly distributed in a $120^\circ \times 30^\circ$ coverage area, and AoAs uniformly distributed in a $120^\circ \times 120^\circ$ coverage area for each of the $k = 1, \cdots, K$ users in the cell.

For the purposes of illustration, we consider simultaneous transmissions to $K = 2$ users. For scheduling, we implement a *directional avoidance* protocol with the dominant cluster in the channel of the first user separated spatially from the dominant cluster in the channel of the second user, as parsed by $\mathcal{F}_{\mathsf{tr}}$ (that is, $\mathbf{f}_{\mathsf{tr}, n_1^2} \neq \mathbf{f}_{\mathsf{tr}, n_1^1}$). The initial beam alignment codebooks are designed based on the beam broadening principles illustrated in [11].

In the first study, we consider the relative performance of the zeroforcing scheme (proposed in Prop. 1) relative to a baseline[2] beam steering scheme with different initial beam alignment codebooks assuming infinite-precision feedback of channel reconstruction parameters and infinite-precision quantization of multi-user beams. Fig. 1(a) illustrates this comparative performance with an $M = 16$ codebook at the user end and two choices of N ($N = 8$ and 32) for different choices of P ($P = 2$ or 4) in approximating $\mathbf{g}_k^\dagger \widehat{\mathbf{H}}_k$ suggesting significant performance improvement over beam steering.

In general, it is intuitive that there should be diminishing performance improvement as P increases (as a choice $P >$

[2]Note that beam steering and the scheme of Prop. 1 can be viewed as implemented with phase shifters alone, and phase + amplitude control, respectively.



(a) (b)

Fig. 1. (a) CDF of sum rates for the proposed zeroforcing scheme and a beam steering scheme with $M = 16$ and $N = 8, 32$ assuming infinite-precision feedback. (b) CDF of sum rates of the multi-user schemes with finite-rate feedback for the same setting as in (a).

$\max_k L_k$ cannot help). However, with different granularities of codebooks in the beam alignment phase, we observe the following (numerical studies not provided due to lack of space). Increasing P when the codebook granularity is already poor (small M and N) does not lead to any performance improvement over that observed with $P = 1$ (beam steering). On the other hand, with a high resolution for $\mathcal{F}_{\mathsf{tr}}$ (large N), even a rank-2 approximation appears to be sufficient to reap most of the performance gains. This is because the performance of the baseline (beam steering) scheme is already quite good and significant relative improvement over it with increasing P has a lower likelihood unless the channel has a large number of similar gain clusters (a low-probability event). When M is large and N is small, the beam steering performance is poor and the channel can be better approximated with the higher codebook resolution of $\mathcal{G}_{\mathsf{tr}}^k$ leading to a sustained performance improvement for even up to $P = 4$.

In the second study, we consider a finite-rate feedback version of the infinite-precision schemes considered in Fig. 1(a). For this, we utilize different quantization functions to quantize the parameters necessary for channel reconstruction. For a phase term θ with a dynamic range of $[0, 2\pi)$ (e.g., $\angle\widehat{\mathsf{s}}_{\mathsf{tr},k,\ell}$ and $\angle\beta_{k,\ell}$), we use a uniform quantizer of the form

$$\mathcal{Q}_B(\theta) = \frac{2\pi}{2^B} \cdot \mathsf{round}\left(\frac{2^B}{2\pi} \cdot \theta\right), \tag{27}$$

where $\mathsf{round}(\cdot)$ stands for a function that rounds off the underlying quantity to the nearest integer. For an amplitude term α with a dynamic range of $[0, 1]$ (e.g., $|\beta_{k,\ell}|$), we use a non-uniform quantizer of the form

$$\mathcal{Q}_B(\alpha) = \frac{\mathsf{round}\left((2^B - 1) \cdot \alpha\right)}{2^B - 1}. \tag{28}$$

For a received SNR term ϱ (in dB) with a theoretically unbounded range (e.g., $10\log_{10}\left(\mathsf{SNR}_{\mathsf{rx},\ell}^{(k)}\right)$), we first cap ϱ to a maximum value of ϱ_{max} and quantize a spread of Δ (in dB) with 2^B quantization levels (denoted as ϱ_i) as follows:

$$\varrho_i = \varrho_{\mathsf{max}} - \frac{\Delta}{2^B - 1} \cdot i, \;\; i = 0, \cdots, 2^B - 1. \tag{29}$$

The quantization of ϱ is given as

$$\mathcal{Q}_B(\varrho) = \varrho_{i^\star} \;\; \mathsf{where} \;\; i^\star = \arg \min_{i=0,\cdots,2^B-1} |\varrho - \varrho_i|. \tag{30}$$

TABLE I

B_{feedback} FOR DIFFERENT CHOICES OF P AND N

	$N = 8$	$N = 32$
$P = 2$	21	25
$P = 4$	45	53

In our study, we assume the following: $B_{\mathsf{SNR}} = 3$ bits with $\varrho_{\mathsf{max}} = 30$ dB and $\Delta = 28$ dB in (29). We also assume that $B_{\mathsf{phase}} = B_{\mathsf{corr,\,amp}} = 3$ bits with an $M = 16$ codebook leading to the feedback overhead as in Table I. While this overhead may appear onerous, similar feedback overheads are currently considered viable in 3GPP 5G-NR design via Type-II feedback. Fig. 1(b) illustrates the performance of finite-rate feedback (relative to infinite-precision feedback) for the $N = 8$ and 32 codebooks for initial beam alignment. From this study, we observe that the proposed joint quantization scheme performs comparable with a scheme that uses infinite precision for all the parameters of interest.

In our third study, reported in Fig. 2, we compare the performance of the proposed zeroforcing scheme with the upper bound established in Theorem 1 using $M = 16$ and $N = 8$ or 32. This study shows that the gap to the upper bound is small (up to 1 bps/Hz) as N increases suggesting the good performance of the proposed scheme. Nevertheless, the performance gap suggests the possible utility of more advanced feedback mechanisms, a topic for future research.

V. Conclusion

This paper motivates a hybrid precoding structure where the analog part of the precoder is *simultaneously* controlled by both phase shifters and gain control. With this background, the focus of this work has been on the development of a feedback mechanism to convey estimates of certain quantities of interest from an initial beam alignment phase to enable the base-station to construct an advanced RF precoding

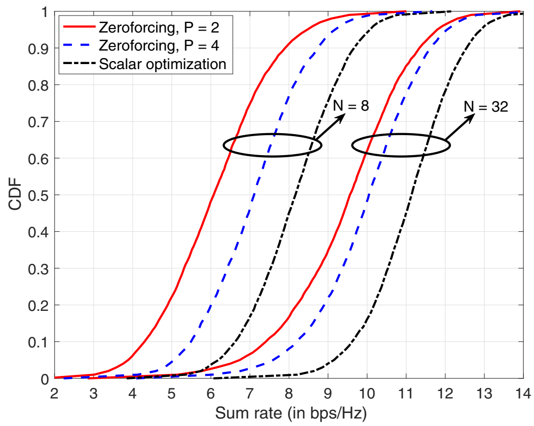


Fig. 2. CDF of sum rates of the multi-user schemes compared with the upper bound of Theorem 1 using a $M = 16$ codebook and $N = 8$ or 32.

structure for multi-user transmissions. These quantities of interest include the top-P (where $P \geq 1$) base-station side beam indices, phases and amplitudes of the received signal estimate, as well as the cross-correlation information of the beams at the user end. This feedback is leveraged to reconstruct/estimate a rank-P approximation of the channel matrix of interest at the base-station end and generate a zeroforcing structure for multi-user interference management. Critical to the implementation of this zeroforcing structure are amplitude and phase control. Numerical studies show that the additional feedback overhead is marginal, but the relative performance improvement over a simplistic beam steering scheme is quite significant even with a very coarse initial beam alignment codebook.

Acknowledgment

This material is based upon work supported in part by the National Science Foundation under grants CCF1403458 and CNS1642982.

References

[1] Aalto University, AT&T, BUPT, CMCC, Ericsson, Huawei, Intel, KT Corporation, Nokia, NTT DOCOMO, NYU, Qualcomm, Samsung, U. Bristol, and USC, "White paper on '5G channel model for bands up to 100 GHz'," *v2.3*, Oct. 2016.

[2] 3GPP TR 38.901 V14.1.1 (2017-07), "Technical Specification Group Radio Access Network; Study on Channel Model for Frequencies from 0.5 to 100 GHz (Rel. 14)," July 2017.

[3] V. Raghavan, A. Partyka, L. Akhoondzadeh-Asl, M. A. Tassoudji, O. H. Koymen, and J. Sanelli, "Millimeter wave channel measurements and implications for PHY layer design," *IEEE Trans. Ant. Propagat.*, vol. 65, no. 12, pp. 6521–6533, Dec. 2017.

[4] V. Raghavan and A. M. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 345–364, Jan. 2011.

[5] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Sig. Proc. Magaz.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[6] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2014.

[7] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Magaz.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[8] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis and measurements," *IEEE Trans. Ant. Propagat.*, vol. 61, no. 7, pp. 3814–3827, July 2013.

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

[10] V. Raghavan, S. Subramanian, J. Cezanne, and A. Sampath, "Directional beamforming for millimeter-wave MIMO systems," *Proc. IEEE Global Telecommun. Conf., San Diego, CA*, pp. 1–7, Dec. 2015.

[11] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. H. Koymen, "Beamforming tradeoffs for initial UE discovery in millimeter-wave MIMO systems," *IEEE Journ. Sel. Topics in Sig. Proc.*, vol. 10, no. 3, pp. 543–559, Apr. 2016.

[12] S. Sun, T. S. Rappaport, R. W. Heath, Jr., A. Nix, and S. Rangan, "MIMO for millimeter wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Magaz.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[13] V. Raghavan, S. Subramanian, J. Cezanne, A. Sampath, O. H. Koymen, and J. Li, "Directional hybrid precoding in millimeter-wave MIMO systems," *Proc. IEEE Global Telecommun. Conf., Washington, DC*, pp. 1–7, Dec. 2016.

[14] ——, "Single-user vs. multi-user precoding for millimeter wave MIMO systems," *IEEE Journ. Sel. Areas in Commun.*, vol. 35, no. 6, pp. 1387–1401, June 2017.

[15] P. Sudarshan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Channel statistics-based joint RF-baseband design for antenna selection for spatial multiplexing," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3501–3511, Dec. 2006.

[16] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Sig. Proc.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[17] A. Adhikary, E. A. Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mm-Wave channels," *IEEE Journ. Sel. Areas in Commun.*, vol. 32, no. 6, pp. 1239–1255, June 2014.

[18] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journ. Sel. Topics in Sig. Proc.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[19] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[20] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journ. Sel. Topics in Sig. Proc.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[21] S. Noh, M. D. Zoltowski, and D. J. Love, "Training sequence design for feedback assisted hybrid beamforming in massive MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 187–200, Jan. 2016.

[22] T. E. Bogale, L. B. Le, A. Haghighat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog-digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.

[23] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE Journ. Sel. Areas in Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[24] H. Krishnaswamy and H. Hashemi, "Integrated beamforming arrays," *In mm-Wave Silicon Technology, (A. M. Niknejad and H. Hashemi, Eds.), Springer, NY*, pp. 243–295, 2008.

[25] G.-L. Huang, S.-G. Zhou, T.-H. Chio, H.-T. Hui, and T.-S. Yeo, "A low profile and low sidelobe wideband slot antenna array fed by an amplitude-tapering waveguide feed-network," *IEEE Trans. Ant. Propagat.*, vol. 63, no. 1, pp. 419–423, Jan. 2015.

[26] Z. Briqech, A.-R. Sebak, and T. A. Denidni, "Low-cost wideband mmWave phased array using the piezoelectric transducer for 5G applications," *IEEE Trans. Ant. Propagat.*, Dec. 2017.

[27] V. Raghavan, A. Partyka, S. Subramanian, A. Sampath, O. H. Koymen, K. Ravid, J. Cezanne, K. K. Mukkavilli, and J. Li, "Millimeter wave MIMO prototype: Measurements and experimental results," *IEEE Commun. Magaz.*, vol. 56, no. 1, pp. 202–209, Jan. 2018.

[28] V. Raghavan, S. V. Hanly, and V. V. Veeravalli, "Statistical beamforming on the Grassmann manifold for the two-user broadcast channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6464–6489, Oct. 2013.

[29] M. R. Castellanos, V. Raghavan, J. H. Ryu, O. H. Koymen, J. Li, D. J. Love, and B. Peleato, "Channel reconstruction-based hybrid precoding for millimeter wave multi-user MIMO systems," *Revised for the IEEE Journ. Sel. Topics in Sig. Proc.*, 2018.