Constant Envelope Precoding for MIMO Systems

Shuowen Zhang, Rui Zhang, and Teng Joon Lim

Abstract

Constant envelope (CE) precoding is an appealing transmission technique, which is realizable with a single radio frequency (RF) chain at the multi-antenna transmitter, while enabling highly efficient power amplification. In this paper, we study the transceiver design for a point-to-point multiple-input multiple-output (MIMO) system with CE precoding. Both single-stream transmission (i.e., beamforming) and multi-stream transmission (spatial multiplexing) are considered. For beamforming mode, we optimize the receive beamforming vector to minimize the symbol error rate (SER) for any given channel realization and desired constellation at the combiner output. By reformulating the problem as an equivalent quadratically constrained quadratic program (QCQP), we propose an efficient semi-definite relaxation (SDR) based algorithm to find an approximate solution. Next, for spatial multiplexing mode, we propose a new scheme based on transmit antenna grouping and receive zero-forcing (ZF) based beamforming, which decomposes the MIMO channel into parallel MIMO sub-channels, each operating with CE precoding in beamforming mode. The transmit antenna grouping and receive beamforming vectors are jointly designed to minimize the maximum SER over all data streams.

This work will be presented in part at the IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, May 23-27, 2016.

S. Zhang is with the NUS Graduate School for Integrative Sciences and Engineering (NGS), National University of Singapore (e-mail:shuowen.zhang@u.nus.edu). She is also with the Department of Electrical and Computer Engineering, National University of Singapore.

R. Zhang is with the Department of Electrical and Computer Engineering, National University of Singapore (e-mail:elezhang@nus.edu.sg). He is also with the Institute for Infocomm Research, A*STAR, Singapore.

T. J. Lim is with the Department of Electrical and Computer Engineering, National University of Singapore (e-mail:eleltj@nus.edu.sg).
Index Terms

Constant envelope (CE) precoding, multiple-input multiple-output (MIMO), beamforming, spatial multiplexing, semi-definite relaxation (SDR).

I. INTRODUCTION

Motivated by the demand for power-efficient and cost-effective radio frequency (RF) components in wireless communication systems, there has been an upsurge of research interests in the so-called constant envelope (CE) precoding in multi-antenna or multiple-input multiple-output (MIMO) communications [1]–[7]. Specifically, under the per-antenna CE constraint that restricts the equivalent complex baseband signal at each transmit antenna to have constant amplitude, CE precoding performs a mapping (which is generally nonlinear) from the desired information-bearing symbols to solely the transmitted signal phases at multi-antennas based on the instantaneous channel state information (CSI).

CE precoding is advantageous compared to its non-CE counterparts due to the following reasons. First, for conventional non-CE precoding techniques, both the amplitude and phase of the equivalent complex baseband signal at each transmit antenna need to vary depending on the instantaneous channel and/or symbol realization. Therefore, they are practically performed in the digital domain and require a dedicated RF chain for each transmit antenna, which is both costly and power-consuming. In contrast, CE precoding can be realized directly in the RF domain by using a network of digitally-controlled analog phase shifters with a single RF chain, as shown in Fig. 1, even when multiple data streams are transmitted concurrently in parallel (i.e., spatial multiplexing). Furthermore, CE precoding leads to much better power amplifier (PA) efficiency, since it feeds the PA input with a CE signal, as shown in Fig. 1. Highly efficient PAs in practice have nonlinear amplitude transfer functions (e.g., class-C and switched mode PAs), hence they can only be used with CE input signals or else output distortion arises [8]. Moreover, since CE signals have the lowest possible peak-to-average power ratio (PAPR), they require the minimum backoff for operation when linear PAs (e.g., class-A and class-B PAs) are used, thus achieving high efficiency [8]. The low PAPR of CE signals also allows for the use of less expensive PAs with smaller dynamic range.
Despite the aforementioned benefits of CE precoding, transceiver design with the per-antenna CE constraint is more challenging than those with the conventional average-based sum power constraint (SPC) and per-antenna power constraint (PAPC) (see e.g., [9]–[13]), which are less restrictive. In [1]–[3], a single-user multiple-input single-output (MISO) system with the per-antenna CE constraint is studied. It is shown in [1], [2] that by varying the transmitted signal phases, the noise-free signal at the receiver always lies in an annular region, whose boundaries are characterized by the instantaneous channel realization and per-antenna transmit power. Moreover, efficient CE precoding algorithms are proposed in [1], [2] to find the nonlinear mapping from any desired received signal point within the annulus to the transmitted signal phases based on the instantaneous CSI. Furthermore, note that a desired receiver constellation is feasible for CE precoding in MISO channel if and only if it can be scaled to lie in the annulus, such that the corresponding transmitted CE signals can be found for all the signal points in the scaled

\[^{1}\text{In practice, a bandpass filter can be applied between each analog phase shifter and transmit antenna to limit the signal bandwidth, which is omitted here for simplicity.}\]

Fig. 1. Transmitter architecture for CE precoding in MIMO system.\[^{1}\]
constellation. Therefore, for a fading channel that yields a time-varying annulus, a fixed receiver constellation may not be always feasible, thus resulting in severe reliability degradation (assuming transmission is blocked when the constellation is not feasible). This problem is successfully tackled in [3], where both fixed-rate and variable-rate adaptive receiver constellation designs are proposed for CE precoding in MISO fading channel.

On the other hand, for multi-user large-scale MISO downlink systems, low-complexity CE precoding algorithms are proposed in [4], [5] for frequency-flat channels and in [6] for frequency-selective channels. It is shown that with a sufficiently large number of transmit antennas, arbitrarily low multi-user interference (MUI) can be achieved at each user with the proposed schemes in [4]–[6]. As an extension to [6], an efficient CE precoding scheme is proposed in [7] considering an additional constraint on the signal phase variation at each transmit antenna between consecutive channel uses, such that the spectral regrowth resulting from abrupt phase changes can be potentially eliminated. It is shown that the proposed scheme in [7] achieves comparable performance with that in [6].

In this paper, we study the transceiver design in a point-to-point MIMO system with CE precoding, assuming perfect CSI is available at both the transmitter and the receiver. Both beamforming mode (with single transmitted data stream) and spatial multiplexing mode (with multiple data streams transmitted concurrently) are considered. Our main contributions are summarized as follows:

- For beamforming mode, we consider the problem of receive beamforming vector optimization to minimize the symbol error rate (SER) at the combiner output, for any given channel realization and desired constellation at the combiner output. Specifically, by approximating the exact SER with its union bound, we formulate the equivalent problem of maximizing the minimum Euclidean distance (MED) between any two signal points at the combiner output while guaranteeing the feasibility of the constellation (i.e., it can be scaled to lie in an annular region characterized by the channel realization and receive beamforming). We first show that this problem is feasible for any desired constellation at the combiner output if the rank of the channel matrix is no smaller than two, which always holds under the assumption of independent and identically distributed (i.i.d.) Rayleigh fading MIMO channel. Then, we introduce an auxiliary vector to reformulate this problem into an equivalent quadratically
constrained quadratic program (QCQP). By applying the semi-definite relaxation (SDR) technique as well as our customized Gaussian randomization methods, we propose an efficient algorithm to find an approximate solution to the QCQP.

- Next, for spatial multiplexing mode, CE precoding that maps the symbols of multiple data streams to the transmitted signal phase at each antenna in general needs to be jointly designed with the MIMO receiver, which is a difficult problem to solve. To tackle this problem, we propose a new scheme with transmit antenna grouping and zero-forcing (ZF) based receive beamforming. As a result, the MIMO channel is decomposed into parallel smaller-size MIMO sub-channels each operating with CE precoding in beamforming mode. We then jointly optimize the receive beamforming vectors for all MIMO sub-channels as well as the transmit antenna grouping to minimize the SER, by maximizing the minimum MED of the received signal constellations over all data streams subject to the constellation feasibility constraints.

The remainder of this paper is organized as follows. Section II introduces the system models for CE precoding in beamforming mode and spatial multiplexing mode, respectively. A new scheme for spatial multiplexing mode is also proposed in Section II. Section III presents the receiver optimization problem for beamforming mode and proposes an efficient algorithm. Section IV studies the transceiver optimization problem for spatial multiplexing mode with the proposed scheme. Numerical results are provided in Section V to evaluate the performance of the proposed schemes. Finally, Section VI concludes the paper.

Notations: Scalars and vectors are denoted by lower-case letters and boldface lower-case letters, respectively. $|z|$, $z^*$, $\arg\{z\}$ and $\mathfrak{Re}\{z\}$ denote the absolute value, the conjugate, the angle and the real part of a complex scalar $z$, respectively. $\|z\|_p$ and $z_k$ denote the $l_p$-norm and the $k$th element of a vector $z$, respectively. $\mathbb{C}^{M \times N}$ denotes the space of $M \times N$ complex matrices. $I_M$ denotes the $M \times M$ identity matrix, and $0$ denotes an all-zero matrix with appropriate dimension if not specified. For an $M \times N$ matrix $A$, $A^T$ and $A^H$ denote its transpose and conjugate transpose, respectively; $\text{rank}(A)$ and $[A]_{i,j}$ denote the rank of $A$ and the $(i,j)$-th element of $A$, respectively. The null space of $A$ is defined as $\text{Null}(A) \triangleq \{x \in \mathbb{C}^{N \times 1} : Ax = 0\}$. For a square matrix $S$, $\text{tr}(S)$ denotes its trace, and $S \succeq 0$ means that $S$ is positive semi-definite. The distribution of a circularly symmetric complex Gaussian (CSCG) random variable with mean $\mu$ and variance $\sigma^2$
is denoted by $CN(\mu, \sigma^2)$; and $\sim$ stands for “distributed as”. $\max\{x, y\}$ and $\min\{x, y\}$ denote the maximum and the minimum of two real numbers $x$ and $y$, respectively.

II. SYSTEM MODEL

Consider a point-to-point MIMO system with $M_t \geq 2$ antennas at the transmitter and $M_r \geq 2$ antennas at the receiver. We assume a quasi-static flat-fading environment with $\tilde{H} \in \mathbb{C}^{M_r \times M_t}$ denoting the equivalent complex baseband channel matrix. The entries of $\tilde{H}$ are modeled by i.i.d. zero-mean CSCG random variables with equal variance of $\beta$, i.e., $[\tilde{H}]_{i,j} \sim CN(0, \beta)$, $\forall i, \forall j$, where $\beta$ specifies the average channel power attenuation due to path loss and shadowing. Note that this implies $\text{rank}(\tilde{H}) = \min\{M_r, M_t\}$, i.e., $\tilde{H}$ is a full-rank matrix, with probability one.

The baseband transmission is modeled by

$$\tilde{y} = \tilde{H}x + \tilde{n},$$

where $\tilde{y} \in \mathbb{C}^{M_r \times 1}$ and $x \in \mathbb{C}^{M_t \times 1}$ denote the received and the transmitted signal vectors, respectively; $\tilde{n} \sim CN(0, \sigma^2 I_{M_r})$ denotes the $M_r \times 1$ CSCG noise vector at the receiver. We consider CE precoding at the transmitter, under the assumption that $\tilde{H}$ is perfectly known at both the transmitter and the receiver. As shown in Fig. 1, we assume a total transmit power denoted by $P$, which is equally allocated to the $M_t$ transmit antennas. With CE precoding, the equivalent complex baseband signal at each transmit antenna is expressed as

$$x_i = \sqrt{\frac{P}{M_t}} e^{j\theta_i}, \quad i = 1, ..., M_t,$$

where information is modulated in the transmitted signal phases $\theta_i \in [0, 2\pi)$, $i = 1, ..., M_t$.

Let $\bar{R}$ denote the transmission rate in bits/second/hertz (bps/Hz). We assume that the transmitted bit sequence is demultiplexed into $K$, $K \leq \min\{M_r, M_t\}$ data streams, each carrying $\frac{\bar{R}}{K}$ bits. For convenience of modulation, we assume $\frac{\bar{R}}{K}$ is an integer. Each data stream is further assumed to be modulated with the same constellation denoted by $\mathcal{S}$, which is of size $N = 2^{\frac{\bar{R}}{K}}$. In the following, we consider the two transmission modes: beamforming (i.e., $K = 1$) and spatial multiplexing (i.e., $K \geq 2$), and present their transceiver models, respectively.
A. Beamforming Mode

For beamforming mode with $K = 1$, we let $u \in \mathbb{C}^{M_t \times 1}$ denote the receive beamforming vector, which is assumed to be normalized such that $\|u\|_2 = 1$ without loss of generality. After applying the receive beamforming, the combiner output signal is given by

$$y = u^H \tilde{y} = u^H \tilde{H} x + n,$$

(3)

where $u^H \tilde{H}$ is the effective MISO channel from the transmitter to the combiner output, and $n = u^H \tilde{n}$ denotes the effective noise, whose distribution can be shown to be given by $n \sim \mathcal{C} \mathcal{N}(0, \sigma^2)$. Let $d \triangleq u^H \tilde{H} x = \sqrt{\frac{P}{M_t}} u^H \tilde{H} [e^{j\theta_1}, ..., e^{j\theta_{M_t}}]^T$ denote the noise-free signal at the combiner output.

First, note that the constellation $\mathcal{S}$ is feasible at the combiner output if and only if there exists a scaling factor $\alpha > 0$, such that any symbol point on $\alpha \mathcal{S}$ can be mapped to CE signals at the transmitter, i.e., the following problem is feasible for any $s \in \mathcal{S}$:

$$\text{find } \{\theta_i\}_{i=1}^{M_t} \quad \text{s.t. } d = \alpha s \quad \theta_i \in [0, 2\pi), \ i = 1, ..., M_t.$$

(4)

By generalizing the results in [1], [2] for the MISO channel, the feasible region of $d$ with a given $u$ and $\theta_i \in [0, 2\pi), \ \forall i$ can be shown to be given by

$$\mathcal{D}(u) = \{d \in \mathbb{C} : r(u) \leq |d| \leq R(u)\},$$

(5)

where

$$R(u) = \sqrt{\frac{P}{M_t}} \|u^H \tilde{H}\|_1,$$

(6)

$$r(u) = \sqrt{\frac{P}{M_t}} \max \left\{2\|u^H \tilde{H}\|_\infty - \|u^H \tilde{H}\|_1, 0\right\}. $$

(7)
As a result of (5), $\mathcal{S}$ is feasible if and only if $\alpha > 0$ exists such that $\alpha \mathcal{S} \subset \mathcal{D}(u)$, or equivalently,

$$
\frac{r(u)}{R(u)} \leq \frac{\min_{s \in \mathcal{S}} |s|}{\max_{s \in \mathcal{S}} |s|}.
$$

Moreover, for any feasible $\mathcal{S}$ and the corresponding $\alpha$, efficient CE precoding algorithms proposed in [1], [2] can be used to find the solution to Problem (4) for any $s \in \mathcal{S}$ based on $u^H \tilde{H}$, for which the details are omitted for brevity. Therefore, the combiner output signal in (3) is equivalently represented by

$$
y = \alpha s + n, \quad s \in \mathcal{S}.
$$

For illustration, the system diagram of CE precoding in beamforming mode is shown in Fig. 2.

**Remark 1:** It is worth noting that in order to maximize the signal power at the combiner output and yet meet the feasibility constraint of $\alpha \mathcal{S} \subset \mathcal{D}(u)$, we should set $\alpha = R(u)$ in (9) for any $\mathcal{S}$ that is feasible and satisfies $\max_{s \in \mathcal{S}} |s| = 1$, such that the signal point with the largest amplitude in $\mathcal{S}$ lies on the outer boundary of $\mathcal{D}(u)$ at the combiner output.

Note that for given channel $\tilde{H}$, $\mathcal{D}(u)$ as well as the feasibility of $\mathcal{S}$ depends on the receive beamforming vector $u$. For example, consider the case where $M_t = M_r = 2$, $|[\tilde{H}]_{1,1}| = $
0.5, $|\tilde{H}_{1,2}| = 0.2$, $|\tilde{H}_{2,1}| = 0.35$, $|\tilde{H}_{2,2}| = 0.25$, and $S$ is a 16-QAM (quadrature amplitude modulation) constellation (i.e., $\min_{s \in S}|s|/\max_{s \in S}|s| = \frac{1}{3}$). As shown in Fig. 3, $S$ is infeasible with $u^{(1)} = [1, 0]^T$, but is feasible with $u^{(2)} = [0, 1]^T$. As a result, we are motivated to investigate the design of $u$ based on the desired constellation $S$ and channel realization $\tilde{H}$, which will be detailed in Section III.

![Diagram](image)

(a) Infeasible case with $u^{(1)} = [1, 0]^T$

(b) Feasible case with $u^{(2)} = [0, 1]^T$

Fig. 3. Feasibility of 16-QAM for CE precoding in beamforming mode with given $\tilde{H}$ and different $u$.

### B. Spatial Multiplexing Mode

For spatial multiplexing mode with $K \geq 2$, we assume a linear receiver is used to decode $s_k$’s. Specifically, let $u_k \in \mathbb{C}^{M_r \times 1}$ denote the receive beamforming vector for decoding $s_k$, which is assumed to be normalized such that $\|u_k\|_2 = 1$ without loss of generality. Applying $u_k^H$ to the received signal vector in (1) yields

$$y_k = u_k^H \tilde{y} = u_k^H \tilde{H} x + n_k,$$

(10)

where $u_k^H \tilde{H}$ is the effective MISO channel from the transmitter to the decoding output of the $k$th data stream, and $n_k = u_k^H \tilde{n}$, with $n_k \sim \mathcal{CN}(0, \sigma^2)$. Let $d_k \triangleq u_k^H \tilde{H} x = \sqrt{\frac{P}{M_t}} u_k^H \tilde{H} \left[ e^{j\theta_1}, ..., e^{j\theta_M} \right]^T$ denote the noise-free signal received from the $k$th data stream. Note that $d_k$’s are coupled with all $\theta_i$'s, which introduces the following challenges to the transceiver design:

- First, note that with given $\{u_k\}_{k=1}^K$, $S$ is feasible for the $K$ data streams if and only if there exists a set of scaling factors $\{\alpha_k : \alpha_k > 0\}_{k=1}^K$, such that the following problem is feasible
for any \( \{s_k : s_k \in \mathcal{S}\}_{k=1}^K \):

\[
\text{find } \{\theta_i\}_{i=1}^{M_t} \\
\text{s.t. } d_k = \alpha_k s_k, \quad k = 1, ..., K \\
\theta_i \in [0, 2\pi), \quad i = 1, ..., M_t.
\]

(11)

However, this condition is in general difficult to verify when \( K > 1 \). Specifically, it is hard to check the feasibility of Problem (11) for given \( \{\alpha_k, s_k\}_{k=1}^K \), since the jointly feasible region for \( \{d_k\}_{k=1}^K \) with \( \theta_i \in [0, 2\pi), \forall i \) is difficult to characterize.\(^2\)

- Second, even assuming \( \mathcal{S} \) is verified to be feasible with given \( \{u_k\}_{k=1}^K \), it is hard to find the mapping from desired \( \{\alpha_k, s_k\}_{k=1}^K \) to the transmitted signal phases \( \theta_i \)'s by solving Problem (11), which is a non-convex problem and is more difficult to solve than Problem (4) for the case of \( K = 1 \).
- Third, note that both the feasibility of \( \mathcal{S} \) and the CE precoding design depend on the receive beamforming vectors \( \{u_k\}_{k=1}^K \). However, due to the lack of effective methods to deal with the above problems, it is difficult to formulate a problem to optimize \( \{u_k\}_{k=1}^K \) directly.

To negotiate these challenges, we propose a scheme that decouples the CE precoding design for the \( K \) data streams, by adopting antenna grouping at the transmitter and ZF-based beamforming at the receiver. Specifically, the transmit antennas are divided into \( K \) groups with equal size \( \frac{M_t}{K} \), each assigned to the transmission of one data stream.\(^3\) Let \( \tilde{H}_k \in \mathbb{C}^{M_r \times \frac{M_t}{K}} \) denote the channel matrix from the transmit antennas in the \( k \)th group to the receiver. For convenience of illustration, we assume the grouping is based on antenna index, i.e., the first group consists of transmit antennas with indices 1 to \( \frac{M_t}{K} \), and so on, which yields \( \tilde{H}_k = [\tilde{h}_{(k-1)\frac{M_t}{K}+1}, ..., \tilde{h}_{k\frac{M_t}{K}}] \), with \( \tilde{h}_i \) denoting the \( i \)th column vector of \( \tilde{H} \).\(^4\) To eliminate the inter-group interference, the

\(^2\)It is worth noting that although the marginally feasible region of each \( d_k \) can be shown to be still an annular region (same as the beamforming mode), \( \{d_k\}_{k=1}^K \) from all \( K \) data streams may not be jointly feasible with each \( d_k \) arbitrarily drawn from its corresponding annular region.

\(^3\)We assume \( \frac{M_t}{K} \) is an integer in the sequel.

\(^4\)Note that the results are directly extendible to other transmit antenna grouping cases, which will be considered later in Section IV.
following ZF constraints are imposed on the receive beamforming vectors:

\[ u_k^H \tilde{H}_{[-k]} = 0, \quad \forall k, \]  

(12)

where \( \tilde{H}_{[-k]} = [\tilde{H}_1, \ldots, \tilde{H}_{k-1}, \tilde{H}_{k+1}, \ldots, \tilde{H}_K] \). Note that the equalities in (12) have non-trivial solutions (i.e., \( u_k \neq 0, \forall k \)) if and only if \( \text{rank} (\tilde{H}_{[-k]}) = \min \left\{ M_r, \frac{M_r(K-1)}{K} \right\} < M_r, \forall k \)

holds. This implies \( M_r \geq \frac{M_r(K-1)}{K} + 1 \) needs to be true, which is thus assumed in the rest of this paper.\(^5\)

The structures of \( u_k \)'s that satisfy (12) can be simplified as follows. Let the singular value decomposition (SVD) of \( \tilde{H}_{[-k]}^H \) be denoted as

\[ \tilde{H}_{[-k]}^H = U_k \Lambda_k V_k^H = U_k \Lambda_k [\tilde{V}_k \tilde{V}_r]^H, \]  

(13)

where \( U_k \in \mathbb{C}^{(K-1)M_t \times (K-1)M_t} \) and \( V_k \in \mathbb{C}^{M_r \times M_r} \) are unitary matrices, i.e., \( U_k^H U_k = I_{(K-1)M_t} \) and \( V_k^H V_k = I_{M_r} \), and \( \Lambda_k = [\Sigma_k \ 0] \in \mathbb{C}^{(K-1)M_t \times (K-1)M_t} \) with \( \Sigma_k \in \mathbb{C}^{(K-1)M_t \times (K-1)M_t} \)

being a diagonal matrix. Furthermore, \( \tilde{V}_k \in \mathbb{C}^{M_r \times (K-1)M_t} \) and \( \tilde{V}_k \in \mathbb{C}^{M_r \times \left( M_r - \frac{(K-1)M_t}{K} \right)} \) consist of the first \( \frac{(K-1)M_t}{K} \) and the last \( M_r - \frac{(K-1)M_t}{K} \) right singular vectors of \( \tilde{H}_{[-k]}^H \), respectively. It can be shown that \( \tilde{V}_k \) with \( \tilde{V}_k^H \tilde{V}_k = I_{M_r - \frac{(K-1)M_t}{K}} \) forms an orthogonal basis for the null space of \( \tilde{H}_{[-k]}^H \). Therefore, to guarantee \( u_k^H \tilde{H}_{[-k]} = 0 \), \( u_k \) must be in the following form:

\[ u_k = \tilde{V}_k w_k, \]  

(14)

where \( w_k \in \mathbb{C}^{M_r - \frac{(K-1)M_t}{K} \times 1} \).

By defining \( \bar{H}_k = \tilde{V}_k^H \tilde{H}_k \), we have \( u_k^H \bar{H}_k = [ \underbrace{0, \ldots, 0}_{1, \ldots, \frac{(K-1)M_t}{K}} \underbrace{w_k^H \bar{H}_k}_{\frac{kM_t}{K} + 1, \ldots, M_t} \underbrace{0, \ldots, 0}_{kM_t} ]. \) (10) is thus rewritten as

\[ y_k = w_k^H \bar{H}_k \bar{x}_k + n_k, \]  

(15)

where \( \bar{x}_k = [\bar{x}_k \frac{(K-1)M_t}{K} + 1, \ldots, \bar{x}_k \frac{kM_t}{K}]^T \) denotes the transmitted signal vector for the \( k \)th group. As a result of (15), the \( K \) data streams are transmitted in parallel over \( K \) MIMO sub-channels \( \bar{H}_k \)'s, each operating in beamforming mode with transmitted signal vector \( \bar{x}_k \) and receive beamforming

\(^5\)It is also worth noting that our results can be extended to the case of \( M_r < \frac{M_r(K-1)}{K} + 1 \), by switching off an appropriate number of transmit antennas.
vector $w_k$. Let $D_k(w_k)$ denote the feasible region of $d_k = w_k^H H_k x_k$, which is similarly defined as (5) for the beamforming mode with $K = 1$. Note that by following similar procedures as in the previous subsection, the feasibilities of $S$ for the $K$ data streams can be separately verified based on $D_k(w_k)$’s; in addition, Problem (11) can now be solved by finding each $x_k$ that yields $d_k = \alpha_k s_k$ separately for all $k$’s. For illustration, the system diagram of our proposed scheme for CE precoding in spatial multiplexing mode is shown in Fig. 4. The transmit antenna grouping and receive beamforming design with our proposed scheme will be addressed in Section IV.

Fig. 4. System diagram of MIMO CE precoding in spatial multiplexing mode with transmit antenna grouping and ZF-based receive beamforming.
III. RECEIVER OPTIMIZATION FOR BEAMFORMING MODE

A. Problem Formulation

For beamforming mode, our objective is to minimize the SER at the combiner output by optimizing the receive beamforming vector $u$ for given $\tilde{H}$ and $S$. Note that since minimizing the exact SER, $P_s$, is in general a difficult problem, we aim to minimize its union bound instead. We assume $S$ is an equiprobable signal set and maximum likelihood (ML) detection is used at the combiner output to recover the signal point in $S$. Without loss of generality, we further assume $\max_{s \in S} |s| = 1$ for the rest of the paper. The union bound of $P_s$ is thus given by

$$P_s \leq (N - 1)Q\left(\sqrt{\frac{(d_{\min}^c)^2}{2\sigma^2}}\right), \quad (16)$$

where $d_{\min}^c = R(u)d_{\min}$ denotes the MED between any two signal points in the scaled constellation $R(u)S$ at the combiner output, with $d_{\min}$ denoting the MED of $S$ [14]. As can be observed from (16), minimizing the union bound of $P_s$ is equivalent to maximizing $d_{\min}^c$, for which we formulate the following optimization problem with given $\tilde{H}$ and $S$ as

$$(P1) \quad \max_u \| u^H \tilde{H} \|_1 \quad (17)$$

subject to

$$\| u \|_2 = 1 \quad (18)$$

$$\max \left\{ 2\| u^H \tilde{H} \|_\infty - \| u^H \tilde{H} \|_1, 0 \right\} \leq \tau, \quad (19)$$

where $\tau = \min_{s \in S} |s| \in [0, 1]$, and the feasibility constraint of $S$ given in (8) is explicitly expressed in (19).

Problem (P1) can be equivalently rewritten as

$$(P2) \quad \max_u \| u^H \tilde{H} \|_1 \quad (20)$$

subject to

$$\| u \|_2 \leq 1 \quad (21)$$

$$\| u^H \tilde{H} \|_\infty \leq \frac{\tau + 1}{2} \| u^H \tilde{H} \|_1 \quad (22)$$

$$\| u^H \tilde{H} \|_1 > 0, \quad (23)$$
since it can be shown that $u^*$ is optimal for Problem (P1) if and only if $u^*$ is the optimal solution to Problem (P2), by noting that the constraint in (21) must be satisfied with equality by the optimal solution to Problem (P2).

Note that Problem (P2) is a non-convex optimization problem since the constraints in (22) and (23) are non-convex. It is also worth noting that Problem (P2) without the constraint in (22) can be shown to be equivalent to the class of unimodular quadratic programs (UQPs) that are known to be NP-hard [15]. Moreover, it is non-trivial to extend the existing approaches for finding approximate solutions to the UQPs (e.g., algorithms based on SDR [15], [16] or fixed-point iterations [17]) to the case of Problem (P2), due to the new non-convex constraint in (22). As a result, Problem (P2) is in general a difficult problem to solve.

In the following, we first study the feasibility of Problem (P2). Then, we provide an efficient algorithm based on SDR to find an approximate solution for this problem.

**B. Feasibility of Problem (P2)**

The feasibility of Problem (P2) can be verified by solving the following problem:

$$\text{(P2-F)} \quad \text{find} \quad u$$

s.t.  

$$\|u^H \tilde{H}\|_\infty \leq \frac{\tau + 1}{2} \|u^H \tilde{H}\|_1$$

$$\|u^H \tilde{H}\|_1 > 0.$$ (26)

Specifically, any feasible solution to Problem (P2) is also a feasible solution to Problem (P2-F); on the other hand, for any feasible solution $u$ to Problem (P2-F), $\frac{u}{\|u\|_2}$ is a feasible solution to Problem (P2). Although Problem (P2-F) is in general difficult to solve due to the non-convex constraints, useful insights can be drawn by investigating its structure, as shown in the following proposition.

**Proposition 1:** Problem (P2) is feasible if $\text{rank}(\tilde{H}) \geq 2$.

**Proof:** Please refer to Appendix A.

Based on Proposition 1, Problem (P2) is always feasible under the assumed i.i.d. Rayleigh fading MIMO channel with $M_t, M_r \geq 2$. Moreover, we provide the following lemma.
Lemma 1: The optimal \( \bar{u} \) to the following problem is a feasible solution to Problem (P2):

\[
(P2-FS) \quad \max_{\bar{u},l} \Re \left\{ \bar{u}^H \sum_{i=1}^{M_t} \hat{h}_i \right\} \\
\text{s.t.} \quad \bar{u}^H \left( 2\hat{h}_l - \sum_{i=1}^{M_t} \hat{h}_i \right) = 0 \tag{27}
\]

\[
\|\bar{u}\|_2 \leq 1 \tag{28}
\]

\[l \in \{1, \ldots, M_t\}. \tag{29}\]

Proof: Please refer to Appendix B.

Based on Lemma 1, a feasible solution to Problem (P2) can be obtained by solving Problem (P2-FS), as summarized in Algorithm 1. It is worth noting that Problem (P2-FS) with given \( l \) is a convex optimization problem, which can be efficiently solved via existing software, e.g., CVX [18].

\begin{algorithm}
\textbf{Algorithm 1:} Algorithm for finding a feasible solution to Problem (P2)

\textbf{Input:} \( \tilde{H} \)
\textbf{Output:} \( \tilde{u}_f \)
\[\text{for } l = 1 \text{ to } M_t \text{ do} \]
\[\quad \text{Obtain } \bar{u}_l^* \text{ and } \bar{v}_l^* \text{ as the optimal solution and optimal value for Problem (P2-FS) with given } l, \text{ respectively.} \]
\[\text{end} \]
\[\text{Obtain } l^* = \arg \max_{l \in \{1, \ldots, M_t\}} \bar{v}_l^*. \]
\[\text{Set } \tilde{u}_f = \bar{u}_{l^*}. \]
\end{algorithm}

C. Proposed Solution to Problem (P2)

First, we introduce an auxiliary vector \( \tilde{p} \in \mathbb{C}^{M_t \times 1} \) with \( |p_i| = 1, \ i = 1, \ldots, M_t \). The objective function of Problem (P2) can be shown to be equivalently given by

\[
\|u^H \tilde{H}\|_1 = \max_{|p_i| = 1, \ i = 1, \ldots, M_t} \Re \left\{ u^H \tilde{H} \tilde{p} \right\}. \tag{31}
\]
For any given \( u \), denote \( p^*(u) \) as the optimal solution to the problem on the right hand side (RHS) of (31), whose elements can be shown to be given by

\[
p^*_i(u) = e^{-j \arg \{ u^H \tilde{h}_i \}}, \quad i = 1, ..., M_t. \tag{32}\]

With (31) and (32), we have the following proposition.

**Proposition 2:** Problem (P2) is equivalent to the following problem:

\[
(P3) \quad \max_{u, p} \mathcal{R} \{ u^H \tilde{H} p \} \tag{33}
\]

\[
\text{s.t.} \quad \|u\|_2 \leq 1 \tag{34}
\]

\[
\|u^H \tilde{H}\|_\infty \leq \frac{\tau + 1}{2} \mathcal{R} \{ u^H \tilde{H} p \} \tag{35}
\]

\[
|p_i| = 1, \quad i = 1, ..., M_t. \tag{36}
\]

**Proof:** Please refer to Appendix C.

Problem (P3) can be shown to be a non-convex QCQP. Next, we propose a customized SDR-based algorithm for solving it. Specifically, we define \( w = [u^T \ p^T]^T \in \mathbb{C}^{(M_r + M_t) \times 1} \), \( W = w w^H \) and formulate the following problem:

\[
(P3-SDR) \quad \max_{W} \mathcal{R} \{ \text{tr}(AWG) \} \tag{37}
\]

\[
\text{s.t.} \quad \text{tr}(W) \leq M_t + 1 \tag{38}
\]

\[
|e_i^T AW g_i| \leq \frac{\tau + 1}{2} \mathcal{R} \{ \text{tr}(AWG) \}, \quad i = 1, ..., M_t \tag{39}
\]

\[
[W]_{i,i} = 1, \quad i = M_r + 1, ..., M_r + M_t \tag{40}
\]

\[
W \succeq 0, \tag{41}
\]

where \( A = [0_{M_t \times M_r}, I_{M_t}] \); \( G = [\tilde{H}^T \ 0_{M_t \times M_t}]^T \) with the \( i \)th column vector denoted by \( g_i \); \( e_i \) denotes the \( i \)th column vector of \( I_{M_t} \). It can be shown that Problem (P3) is equivalent to Problem (P3-SDR) with the additional constraint of \( \text{rank}(W) = 1 \). Therefore, the optimal value of Problem (P3-SDR) is in general an upper bound on those of Problem (P3) and Problem (P2).

Problem (P3-SDR) is a semi-definite programming (SDP), which can be efficiently solved via existing software, e.g. CVX [18]. Let \( W^* \) and \((u^*, p^*)\) denote the optimal solutions to
Problem (P3-SDR) and Problem (P3), respectively. If \( \text{rank}(W^*) = 1 \), our relaxation is tight and \( w^* = [u^T \ p^*]^T \) can be obtained from the eigenvalue decomposition (EVD) of \( W^* \). The optimal solution to Problem (P2) is thus obtained as \( u^* \). Otherwise, for the case of \( \text{rank}(W^*) > 1 \), we aim to extract an approximate solution to Problem (P2) from \( W^* \), for which a commonly adopted approach is via the so-called Gaussian randomization method (see e.g., [16] and references therein). By customizing this method to our problem, we propose two randomization algorithms denoted by \( \text{Rand}_u \) and \( \text{Rand}_p \), which are summarized as Algorithm 2 and Algorithm 3, respectively:

- \( \text{Rand}_u \): Randomly generate \( L_u \ u \)’s based on \( W^*_u \in \mathbb{C}^{M_r \times M_r} \), which is given by \( [W^*_u]_{i,j} = [W^*]_{i,j} \). Choose the solution that satisfies the constraints in (21), (22) and (23), and also yields the largest objective value of Problem (P2).

**Algorithm 2:** \( \text{Rand}_u \)

**Input:** \( W^* \), \( \tilde{H} \), \( \tau \), \( L_u \)

**Output:** \( \tilde{u}_u \)

Obtain \( W^*_u \in \mathbb{C}^{M_r \times M_r} \) by \( [W^*_u]_{i,j} = [W^*]_{i,j} \), \( i = 1, \ldots, M_r \), \( j = 1, \ldots, M_r \).

if \( \text{rank}(W^*_u) = 1 \) then

Obtain \( u_u \) by \( W^*_u = \tilde{u}_u \tilde{u}_u^H \).

if \( \tilde{u}_u \) does not satisfy (22) or (23) then

\( \tilde{u}_u = 0 \).

end

else

Obtain the Cholesky decomposition of \( W^*_u \) by \( W^*_u = V_u V_u^H \).

for \( l = 1 \) to \( L_u \) do

Generate \( r^{(l)} \sim \mathcal{CN}(0, I_{M_r}) \). Obtain \( \tilde{u}^{(l)} = \frac{V_u r^{(l)}}{\|V_u r^{(l)}\|_2} \).

if \( \tilde{u}^{(l)} \) does not satisfy (22) or (23) then

\( \tilde{u}^{(l)} = 0 \).

end

Set \( l^* = \arg \max_{l=1, \ldots, L_u} \|\tilde{u}^{(l)} H\|_1 \), \( \tilde{u}_u = \tilde{u}^{(l^*)} \).

end

- \( \text{Rand}_p \): Randomly generate \( L_p \ p \)’s that satisfy the constraint in (36) based on \( W^*_p \in \mathbb{C}^{M_t \times M_t} \), which is given by \( [W^*_p]_{i,j} = [W^*]_{i+M_t, j+M_t} \). Solve Problem (P3) with each generated \( p \), and choose the resulting optimal \( u \) that yields the largest objective value.
of Problem (P2).

Algorithm 3: \texttt{Rand}_p

\begin{algorithm}
\textbf{Input:} \( W^*, \tilde{H}, \tau, L_p \)  \\
\textbf{Output:} \( \tilde{u}_p \)  \\
Obtain \( W^*_p \in \mathbb{C}^{M_t \times M_t} \) by \( [W^*_p]_{i,j} = [W^*]_{M_t+i,M_t+j}, \ i = 1, \ldots, M_t, \ j = 1, \ldots, M_t. \)  \\
\textbf{if} rank\((W^*_p) = 1\) \textbf{then}  \\
\quad Obtain \( \tilde{p} \) by \( W^*_p = \tilde{p}\tilde{p}^H. \)  \\
\quad \textbf{if \ Problem (P3) is infeasible with given} \ p = \tilde{p} \textbf{then}  \\
\qquad \tilde{u}_p = 0.  \\
\quad \textbf{else}  \\
\qquad Obtain \( \tilde{u}_p \) as the optimal solution to Problem (P3) with given \( p = \tilde{p}. \)  \\
\textbf{end}  \\
\textbf{else}  \\
\quad Obtain the Cholesky decomposition of \( W^*_p \) by \( W^*_p = V_p V_p^H. \)  \\
\quad \textbf{for} \ l = 1 \ \textbf{to} \ L_p \ \textbf{do}  \\
\qquad Generate \( r^{(l)} \sim \mathcal{CN}(0, I_{M_t}). \) Obtain \( \xi^{(l)} = V_p r^{(l)}. \)  \\
\qquad Obtain \( \tilde{p}^{(l)} = [\tilde{p}_1^{(l)}, \ldots, \tilde{p}_{M_t}^{(l)}]^T, \ \text{by \ } \tilde{p}_i^{(l)} = e^{j \arg\{\xi_i^{(l)}\}}, \ i = 1, \ldots, M_t. \)  \\
\qquad \textbf{if \ Problem (P3) is infeasible with given} \ p = \tilde{p}^{(l)} \textbf{then}  \\
\qquad\qquad \tilde{u}^{(l)} = 0.  \\
\qquad \textbf{else}  \\
\quad\quad Obtain \( \tilde{u}^{(l)} \) as the optimal solution to Problem (P3) with given \( p = \tilde{p}^{(l)}. \)  \\
\quad\textbf{end}  \\
\textbf{end}  \\
\quad \text{Set} \ l^* = \arg \max_{l=1,\ldots,L_p} \|\tilde{u}^{(l)H}\tilde{H}\|_1, \ \tilde{u}_p = \tilde{u}^{(l^*)}.  \\
\textbf{end}  \\
\end{algorithm}

However, it is worth noting that due to the non-convex constraint given in (22) of Problem (P2), the feasibility of the approximate solution obtained by Algorithm 2 or Algorithm 3 cannot be guaranteed in general (i.e., \( \tilde{u}_u = 0 \) or \( \tilde{u}_p = 0 \) may occur). Therefore, we propose to employ both Algorithm 2 and Algorithm 3 to find \( \tilde{u}_u \) and \( \tilde{u}_p, \) respectively, based on \( W^*; \) while we also use Algorithm 1 to find a feasible solution \( \tilde{u}_f. \) Then, the approximate solution to Problem (P2) is chosen from \( \tilde{u}_u, \tilde{u}_p \) and \( \tilde{u}_f \) as the one that achieves the maximum objective value of Problem (P2). It is worth noting that since \( \tilde{u}_f \) is always a feasible solution to Problem (P2), the feasibility of the selected solution is guaranteed.

To summarize, we provide Algorithm 4, which finds an approximate solution to Problem (P2) denoted by \( \tilde{u}. \) Note that \( \tilde{u} \) is always feasible for Problem (P2), and is optimal if rank\((W^*) = 1.\)
Algorithm 4: Algorithm for finding an approximate solution to Problem (P2)

\begin{itemize}
\item \textbf{Input:} $\tilde{H}$, $\tau$, $L_u$, $L_p$
\item \textbf{Output:} $\tilde{u}$
\end{itemize}

Obtain $W^*$ by solving Problem (P3-SDR).

\begin{itemize}
\item if $\text{rank}(W^*) = 1$ then
  \begin{itemize}
  \item Obtain $w^*$ by $W^* = w^*w^{*H}$.
  \item Obtain $\tilde{u} = u^*$ by $u^*_j = \tilde{u}^*_j$, $j = 1, ..., M_r$.
  \end{itemize}
\item else
  \begin{itemize}
  \item Obtain $\tilde{u}_f$ via Algorithm 1. Obtain $\tilde{u}_u$ via Algorithm 2. Obtain $\tilde{u}_p$ via Algorithm 3.
  \item Obtain $\tilde{u} = \arg \max_{u_f, u_u, u_p} \left\{ \|\tilde{u}_f^H \tilde{H}\|_1, \|\tilde{u}_u^H \tilde{H}\|_1, \|\tilde{u}_p^H \tilde{H}\|_1 \right\}$.
  \end{itemize}
\end{itemize}

IV. TRANSCEIVER OPTIMIZATION FOR SPATIAL MULTIPLEXING MODE

For spatial multiplexing mode, our objective is to minimize the maximum SER over the $K$ data streams, by jointly optimizing $\{w_k\}_{k=1}^K$ and the transmit antenna grouping with given $\tilde{H}$ and $S$. Note that similar to the beamforming mode, by approximating the exact SERs of the $K$ data streams with their union bounds, this is equivalent to maximizing the minimum MED of the received signal constellations over all $K$ data streams subject to the feasibility constraints of the constellations.

First, we consider the optimization of $\{w_k\}_{k=1}^K$ with fixed transmit antenna grouping. Recall that with our proposed scheme, the $K$ data streams are transmitted in $K$ parallel MIMO sub-channels, each operating with CE precoding in beamforming mode. Therefore, the aforementioned problem in this case is equivalent to maximizing the MED for each of the $K$ data streams separately, by solving the following problem for every $k \in \{1, ..., K\}$:

\begin{equation}
\text{(P4)} \quad \max_{w_k} \|w_k^H H_k\|_1 \quad (42)
\end{equation}

\begin{equation}
\text{s.t.} \quad \|w_k\|_2 \leq 1 \quad (43)
\end{equation}

\begin{equation}
\|w_k^H H_k\|_{\infty} \leq \frac{\tau + 1}{2} \|w_k^H H_k\|_1 \quad (44)
\end{equation}

\begin{equation}
\|w_k^H H_k\|_1 > 0. \quad (45)
\end{equation}

Notice that Problem (P4) is in the same form as Problem (P2). Therefore, the solution to Problem
(P4) can be readily obtained by applying Algorithm 4.

Next, we investigate the feasibility of our proposed scheme based on Problem (P4). Clearly, our proposed scheme is feasible if and only if there exists a transmit antenna grouping, such that Problem (P4) is feasible for any \( k \). By generalizing the result in Proposition 1, we provide a sufficient condition under which the feasibility of our proposed spatial multiplexing scheme is guaranteed, as shown in the following proposition.

**Proposition 3:** Problem (P4) is feasible for all \( k \in \{1, \ldots, K\} \) with any given transmit antenna grouping, if \( M_r \geq \frac{(K-1)M_t}{K} + 2 \) and \( \frac{M_t}{K} \geq 2 \).

**Proof:** Please refer to Appendix D.

It is worth noting that for the case of \( 2K \leq M_r < \frac{(K-1)M_t}{K} + 2 \) and \( \frac{M_t}{K} \geq 2 \), our proposed scheme can still be made feasible by selecting a subset of \( M'_t, M'_t \in \left[ 2K, \frac{K(M_r-2)}{K-1} \right] \) transmit antennas for CE precoding with the other antennas not used.

Finally, by solving Problem (P4) for all \( k \in \{1, \ldots, K\} \) with \( \{H_k\}_{k=1}^K \) resulting from all possible transmit antenna groupings, the optimal grouping can be obtained as the one that yields the maximum minimum MED over all data streams.

### V. Numerical Results

In this section, we provide numerical results to corroborate our study. We assume \( \beta = -90 \)dB. The average noise power at each receive antenna is set to be \( \sigma^2 = -94 \)dBm. The average signal-to-noise ratio (SNR) is defined as \( \text{SNR} = \frac{P\beta}{\sigma^2} \). We further assume \( S \) is an \( N \)-ary QAM constellation unless specified otherwise. The numbers of randomization trials for Algorithm 2 and Algorithm 3 are set as \( L_u = 50 \) and \( L_p = 50 \), respectively.

#### A. Beamforming Mode

In this subsection, we consider the case of \( K = 1 \) and evaluate the performance of the proposed scheme for CE precoding in beamforming mode.

1) **Performance of Algorithm 4:** First, note that the approximate solution \( \tilde{u} \) to Problem (P2) obtained by Algorithm 4 is optimal (i.e., \( \tilde{u} = u^* \)) if \( \text{rank}(W^*) = 1 \), and suboptimal in general if obtained by one of Algorithms 1, 2 and 3 otherwise. In Table I, we show the percentage of
occurrence for the four possible outcomes of \( \tilde{u} \) over \( 10^4 \) independent channel realizations under various setups of \( M_t \) and \( M_r \), with \( N = 16 \) or \( N = 64 \).

For all setups, it is observed from Table I that the percentage of \( \tilde{u} = u^* \) is very large, which shows that it is most likely that the SDR is tight and Algorithm 4 finds the optimal solution to Problem (P2). On the other hand, it is observed that the chance of \( \tilde{u} = \tilde{u}_f \) is zero under all setups. This suggests that using the two Gaussian randomization methods given in Algorithm 2 and Algorithm 3 with the given numbers of randomization trials suffices for finding a feasible solution to Problem (P2), which is generally better than that obtained by Algorithm 1. It is also observed that the solution in the case of \( \text{rank}(W^*) > 1 \) can be either \( \tilde{u} = \tilde{u}_u \) or \( \tilde{u} = \tilde{u}_p \) in general, which shows that employing Algorithm 2 and Algorithm 3 jointly helps improve the performance.

| \( N \) | \( M_t \) | \( M_r \) | \( \tilde{u} = u^* \) | \( \tilde{u} = \tilde{u}_f \) | \( \tilde{u} = \tilde{u}_u \) | \( \tilde{u} = \tilde{u}_p \) |
|---|---|---|---|---|---|---|
| 16 | 2 | 4 | 99.71% | 0.00% | 0.00% | 0.29% |
| 16 | 4 | 2 | 99.53% | 0.00% | 0.12% | 0.35% |
| 16 | 4 | 4 | 99.01% | 0.00% | 0.22% | 0.77% |
| 64 | 2 | 4 | 96.46% | 0.00% | 0.02% | 3.52% |
| 64 | 4 | 2 | 99.25% | 0.00% | 0.11% | 0.64% |
| 64 | 4 | 4 | 98.94% | 0.00% | 0.09% | 0.97% |

2) Performance Comparison of Receive Beamforming Schemes: Next, we compare the performance of the proposed receive beamforming scheme with the following benchmark schemes:

- **Antenna Selection (AS):** In this scheme, the \( j \)th element in the receive beamforming vector is given by \( u_j = 1 \) if \( j = j^* \), and \( u_j = 0 \) otherwise, where \( j^* \) denotes the optimal solution to the following problem:

\[
\max_{j=1,\ldots,M_r} \| \tilde{h}_j' \|_1
\]

s.t. \( \| \tilde{h}_j' \|_\infty \leq \frac{\tau + 1}{2} \| \tilde{h}_j' \|_1 \),

where \( \tilde{h}_j' \) denotes the transposed vector of the \( j \)th row of \( \tilde{H} \). Problem (46) can be easily solved via one-dimensional search over \( j \). If Problem (46) is infeasible, we set \( j^* = \)
arg max_{j=1,\ldots,M_r} \|\tilde{h}'_j\|_1.

- **Strongest Eigenmode Beamforming (SEB):** In this scheme, the receive beamforming vector is obtained as the optimal solution to the following problem:

\[
\max_{\|\mathbf{u}\|_2 \leq 1} \|\mathbf{u}^H \tilde{H} \mathbf{H}^H\|_2,
\]

which can be shown to be the eigenvector corresponding to the maximum eigenvalue of $\tilde{H} \mathbf{H}^H$.

In Fig. 5, we consider the case of $N = 16$ and show the average SERs of our proposed scheme and the benchmark schemes under the following setups: i) $M_t = 2, M_r = 4$, ii) $M_t = 4, M_r = 2$, and iii) $M_t = M_r = 4$. The results are averaged over $10^6$ independent channel realizations. Note that for AS or SEB with 16-QAM constellation (i.e., $\min_{s \in S} |s| = \tau = \frac{1}{3}$), high SER can occur if the resulting receive beamforming vector $\mathbf{u}$ does not satisfy the constraint in (22), thus is infeasible for CE precoding. Therefore, we also show in Fig. 5 the average SER of AS and SEB with hybrid 16-QAM/16-PSK (phase shift keying) constellations, where the constellation $S$ at the combiner output is adaptively switched to 16-PSK if AS or SEB is infeasible with 16-QAM, to achieve the same transmission rate. Note that such schemes are always feasible, since 16-PSK constellation yields $\min_{s \in S} |s| = \tau = 1$, thus the constraint in (22) is always satisfied.

For all three setups, it is observed from Fig. 5 that our proposed scheme outperforms both AS and SEB with 16-QAM. Specifically, AS with 16-QAM results in error floor for the case of $M_t = 2, M_r = 4$, and has an SNR loss of 1.73dB and 2.52dB compared with our proposed scheme at the average SER of $10^{-4}$ for the case of $M_t = 4, M_r = 2$ and $M_t = M_r = 4$, respectively. On the other hand, SEB with 16-QAM results in error floor under all three setups. Note that the performance gain of our proposed scheme is due to the optimization of $\mathbf{u}$, as well as the fact that AS and SEB with 16-QAM may not be always feasible for CE precoding with any channel realization, while our proposed scheme is always feasible (as a consequence of Proposition 1). Moreover, it is observed that our proposed scheme outperforms AS and SEB even with hybrid 16-QAM/16-PSK under all three setups. This implies that compared to using adaptive receiver constellation which requires additional implementation complexity, our proposed design of $\mathbf{u}$ is a more cost-effective method for guaranteeing the feasibility of CE precoding in beamforming.
mode.

Fig. 5. Average SER comparison of receive beamforming schemes.
B. Spatial Multiplexing Mode

In this subsection, we compare the performance of the proposed scheme for CE precoding in spatial multiplexing mode with that in beamforming mode, given the same transmission rate $\bar{R}$. In Fig. 6, we consider $M_t = M_r = 4$ and show the average bit error rate (BER) comparison of the two schemes for the case of $\bar{R} = 2$bps/Hz and $\bar{R} = 8$bps/Hz, respectively. For spatial multiplexing mode, we assume $K = 2$, which is feasible according to Proposition 3. The results are averaged over $10^6$ independent channel realizations.

![Graph](image)

(a) $\bar{R} = 2$bps/Hz

(b) $\bar{R} = 8$bps/Hz

Fig. 6. Average BER comparison of the proposed schemes for CE precoding in beamforming mode and spatial multiplexing mode.

It is observed from Fig. 6 that for both transmission rates, the proposed beamforming scheme outperforms the proposed spatial multiplexing scheme in the high-SNR regime, since it extracts
larger diversity gain from the MIMO channel. Furthermore, it is observed that at the average BER of $10^{-3}$, the proposed beamforming scheme has an SNR gain of $5.14$ dB over the proposed spatial multiplexing scheme for the case of $\bar{R} = 2$ bps/Hz, but suffers from an SNR loss of $4.07$ dB for the case of $\bar{R} = 8$ bps/Hz. This reveals that by exploiting the multiplexing gain of the MIMO channel, the proposed spatial multiplexing scheme is more favorable for CE precoding in the high-rate regime when the SNR is not large.

VI. CONCLUSIONS

This paper investigated the transceiver design for an i.i.d. Rayleigh fading MIMO channel with CE precoding. For beamforming mode, we studied the receive beamforming optimization problem for any channel realization and desired constellation at the combiner output, to maximize the MED between any two signal points at the combiner output subject to the feasibility constraint of the constellation. We showed that this problem is always feasible, and proposed an efficient algorithm based on SDR to find an approximate solution. The proposed receive beamforming scheme was shown to significantly outperform other benchmark schemes in terms of average SER. For spatial multiplexing mode, a new scheme based on transmit antenna grouping and receive ZF-based beamforming was proposed, whose design was further optimized to maximize the minimum MED of the received signal constellations over all data streams subject to the constellation feasibility constraints. Numerical results showed that for fixed transmission rate, the proposed spatial multiplexing scheme achieves better BER performance compared with the proposed beamforming scheme in the high-rate and moderate-SNR regime.

APPENDIX A

PROOF OF PROPOSITION 1

To start, we present the following lemma.

**Lemma 2:** If $\text{rank}(\bar{H}) \geq 2$, there exists $\bar{u}_l \in \mathbb{C}^{M_r \times 1}$ for any $l \in \{1, ..., M_t\}$ that satisfies the following conditions:

$$\bar{u}_l^H \left( 2\bar{h}_l - \sum_{i=1}^{M_t} \bar{h}_i \right) = 0$$

$$\|\bar{u}_l^H \bar{H}\|_1 > 0.$$
Proof: We prove Lemma 2 by contradiction. Suppose any solution \( \bar{u}_l \) to (48) is also a solution to \( \bar{u}_l^H \tilde{H} = 0 \), then it can be shown that \( \text{Null} \left( 2\tilde{h}_l^H - \sum_{i=1}^{M_l} \tilde{h}_i^H \right) \subseteq \text{Null} \left( \tilde{H}^H \right) \). It thus follows that \( \text{rank} \left( \tilde{H}^H \right) \leq \text{rank} \left( 2\tilde{h}_l^H - \sum_{i=1}^{M_l} \tilde{h}_i^H \right) = 1 \). This contradicts the assumption of \( \text{rank} (\tilde{H}) \geq 2 \). The proof of Lemma 2 is thus completed.

With Lemma 2, we prove Proposition 1 by showing that for any \( l \in \{1, \ldots, M_t\} \), any \( \bar{u}_l \) that satisfies (48) and (49) is a feasible solution to Problem (P2-F). First, notice that \( \bar{u}_l \) satisfies the constraint in (26). Then, we show that \( \bar{u}_l \) also satisfies the constraint in (25). Specifically, we have

\[
|\bar{u}_l^H \tilde{h}_i| \overset{(a_1)}{=} |\bar{u}_l^H \sum_{i \neq l} \tilde{h}_i| \overset{(b_1)}{\leq} \sum_{i \neq l} |\bar{u}_l^H \tilde{h}_i| \quad (50)
\]

\[
|\bar{u}_l^H \tilde{h}_j| \overset{(a_2)}{=} |\bar{u}_l^H \tilde{h}_l - \sum_{i \neq l, i \neq j} \bar{u}_l^H \tilde{h}_i| \overset{(b_2)}{\leq} \sum_{i \neq j} |\bar{u}_l^H \tilde{h}_i|, \quad \forall j \in \{1, \ldots, M_t\} \setminus \{l\} \quad (51)
\]

where \((a_1)\) and \((a_2)\) result from (48); \((b_1)\) and \((b_2)\) are due to the triangle inequality. It then follows from (50) and (51) that

\[
|\bar{u}_l^H \tilde{h}_i| \leq \|\bar{u}_l^H \tilde{H}\|_1 - |\bar{u}_l^H \tilde{h}_i|, \quad \forall i \in \{1, \ldots, M_t\},
\]

namely,

\[
\|\bar{u}_l^H \tilde{H}\|_{\infty} \leq \frac{1}{2} \|\bar{u}_l^H \tilde{H}\|_1 \overset{(c)}{\leq} \frac{\tau + 1}{2} \|\bar{u}_l^H \tilde{H}\|_1, \quad (53)
\]

where \((c)\) holds since \( \tau \geq 0 \). Therefore, \( \bar{u}_l \) satisfies the constraint in (25). The proof of Proposition 1 is thus completed.

**APPENDIX B**

**PROOF OF LEMMA 1**

We first show that to prove Lemma 1, it suffices to show that the optimal \( \bar{u} \) to Problem (P2-FS) denoted by \( \bar{u}^\ast \) satisfies \( \|\bar{u}^\ast H\|_1 > 0 \). Specifically, if \( \|\bar{u}^\ast H\|_1 > 0 \), \( \bar{u}^\ast \) can be shown to be a feasible solution to Problem (P2-F) according to the proof of Proposition 1, by noting that it satisfies the constraint in (28); moreover, since \( \bar{u}^\ast \) satisfies the constraint in (29), it is a feasible solution to Problem (P2).
Then, we prove $\|\bar{\mathbf{u}}^* \bar{H}\|_1 > 0$ by contradiction. Suppose, on the contrary, $\|\bar{\mathbf{u}}^* \bar{H}\|_1 = 0$ holds. By noting that $\|\bar{\mathbf{u}}^* \bar{H}\|_1 \geq \Re \left\{ \bar{\mathbf{u}}^* \hat{H} \right\}$ holds for any $\bar{\mathbf{u}}$, it follows that the optimal value of Problem (P2-FS) for any given $l$ is zero, which implies $\text{Null} \left( 2\hat{h}_l^H - \sum_{i=1}^{M_t} \hat{h}_i^H \right) \subseteq \text{Null} \left( \sum_{i=1}^{M_t} \hat{h}_i^H \right)$, $\forall l$. However, note that this is only true if there exist $\beta_l$’s such that $\hat{h}_l = \beta_l \sum_{i=1}^{M_t} \hat{h}_i^H$ holds, i.e., $\text{rank}(\hat{H}) = 1$. This contradicts our assumption of $\text{rank}(\hat{H}) \geq 2$.

The proof of Lemma 1 is thus completed.

APPENDIX C

PROOF OF PROPOSITION 2

First, given any feasible solution $\mathbf{u}$ to Problem (P2), it follows from (31) and (32) that $(\mathbf{u}, \mathbf{p}^*(\mathbf{u}))$ is feasible for Problem (P3) and achieves the same objective value as that of Problem (P2), thus the optimal value of Problem (P3) is no smaller than that of Problem (P2). On the other hand, it follows from (32) that the objective value of Problem (P3) with any feasible solution $(\mathbf{u}, \mathbf{p})$ is always no larger than that with the solution $(\mathbf{u}, \mathbf{p}^*(\mathbf{u}))$. Moreover, note that based on Lemma 1 and (32), $(\bar{\mathbf{u}}^*, \mathbf{p}^*(\bar{\mathbf{u}}^*))$ is a feasible solution to Problem (P3) and yields $\Re \left\{ \bar{\mathbf{u}}^* \hat{H} \mathbf{p}^*(\bar{\mathbf{u}}^*) \right\} > 0$, where $\bar{\mathbf{u}}^*$ denotes the optimal $\bar{\mathbf{u}}$ to Problem (P2-FS). It then follows that the optimal solution to Problem (P3) denoted by $(\mathbf{u}^*, \mathbf{p}^*(\mathbf{u}^*))$ satisfies $\|\mathbf{u}^* \hat{H}\|_1 = \Re \left\{ \mathbf{u}^* \hat{H} \mathbf{p}^*(\mathbf{u}^*) \right\} > 0$, thus $\mathbf{u}^*$ is feasible for Problem (P2) and achieves the same objective value as that of Problem (P3) with the optimal solution. Hence, the optimal value of Problem (P2) is no smaller than that of Problem (P3). Therefore, Problems (P2) and (P3) have the same optimal value. This thus completes the proof of Proposition 2.

APPENDIX D

PROOF OF PROPOSITION 3

We prove Proposition 3 by showing that for any transmit antenna grouping, $M_r \geq \frac{(K-1)M_t}{K} + 2$ and $\frac{M_t}{K} \geq 2$ yield $\text{rank}(\hat{H}_k) \geq 2$, $\forall k$, thus guaranteeing the feasibility of Problem (P4) for all $k$’s according to Proposition 1.

First, note that $\text{rank}(\hat{V}_k) = M_r - \text{rank}(\hat{H}_{[-k]}) = M_r - \frac{(K-1)M_t}{K}$. Then, by noting that $\hat{V}_k^H \hat{H}_{[-k]} = 0$, it can be shown that $\text{rank}(\hat{H}_k) = \text{rank} \left( \hat{V}_k^H \hat{H}_k \right) = \text{rank} \left( \hat{V}_k^H \hat{H} \right) \geq \text{rank} \left( \hat{V}_k^H \right) + \text{rank}(\hat{H}) - M_r = \min\{M_r, M_t\} - \frac{(K-1)M_t}{K}$. Hence, $M_r \geq \frac{(K-1)M_t}{K} + 2$ and $\frac{M_t}{K} \geq 2$ suffice to ensure $\text{rank}(\hat{H}_k) \geq 2$. The proof of Proposition 3 is thus completed.
REFERENCES

[1] S. K. Mohammed and E. G. Larsson, “Single-user beamforming in large-scale MISO systems with per-antenna constant-envelope constraints: the doughnut channel,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3992–4005, Nov. 2012.

[2] J. Pan and W.-K. Ma, “Constant envelope precoding for single-user large-scale MISO channels: efficient precoding and optimal designs,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 982–995, Oct. 2014.

[3] S. Zhang, R. Zhang, and T. J. Lim, “Constant envelope precoding with adaptive receiver constellation in fading channel,” [Online]. Available: http://arxiv.org/abs/1503.09178.

[4] S. K. Mohammed and E. G. Larsson, “Per-antenna constant envelope precoding for large multi-user MIMO systems,” *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1059–1071, Mar. 2013.

[5] J.-C. Chen, C.-K. Wen, and K.-K. Wong, “Improved constant envelope multiuser precoding for massive MIMO systems,” *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1311–1314, Aug. 2014.

[6] S. K. Mohammed and E. G. Larsson, “Constant-envelope multi-user precoding for frequency-selective massive MIMO systems,” *IEEE Wireless Commun. Lett.*, vol. 2, no. 5, pp. 547–550, Oct. 2013.

[7] S. Mukherjee and S. K. Mohammed, “Constant envelope precoding with time-variation constraint on the transmitted phase angles,” *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 221–224, Apr. 2015.

[8] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*. Artech Publishing House, 1999.

[9] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), “Linear precoding via conic optimization for fixed MIMO receivers,” *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[10] W. Yu and T. Lan, “Transmitter optimization for the multi-antenna downlink with per-antenna power constraints,” *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[11] L. Zhang, R. Zhang, Y.-C. Liang, Y. Xin, and H. V. Poor, “On Gaussian MIMO BC-MAC duality with multiple transmit covariance constraints,” *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2064–2078, Apr. 2012.

[12] R. Zhang, “Cooperative multi-cell block diagonalization with per-base-station power constraints,” *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, Dec. 2010.

[13] S. Zhang, R. Zhang, and T. J. Lim, “Massive MIMO with per-antenna power constraint,” in *Proc. IEEE Global Conf. on Signal and Inf. Process. (GlobalSIP)*, Dec. 2014, pp. 642–646.

[14] J. G. Proakis and M. Salehi, *Digital Communications*. McGraw-Hill, 2008.

[15] S. Zhang and Y. Huang, “Complex quadratic optimization and semidefinite programming,” *SIAM J. Optim.*, vol. 16, no. 3, pp. 871–890, 2006.

[16] Z.-Q. Luo, W.-K. Ma, A.-C. So, Y. Ye, and S. Zhang, “Semidefinite relaxation of quadratic optimization problems,” *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[17] M. Soltanalian and P. Stoica, “Designing unimodular codes via quadratic optimization,” *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1221–1234, Mar. 2014.

[18] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming,” version 2.1. [Online] Available: http://cvxr.com/cvx/, Jun. 2015.