Exploiting Spatial Channel Covariance for Hybrid Precoding in Massive MIMO Systems

Sungwoo Park, *Student Member, IEEE*, Jeonghun Park, *Student Member, IEEE*, Ali Yazdan, *Student Member, IEEE*, and Robert W. Heath, Jr., *Fellow, IEEE*

***Abstract*—We propose a new hybrid precoding technique for massive multi-input multi-output (MIMO) systems using spatial channel covariance matrices in the analog precoder design. Applying a regularized zero-forcing precoder for the baseband precoding matrix, we find an unconstrained analog precoder that maximizes signal-to-leakage-plus-noise ratio (SLNR) while ignoring analog phase shifter constraints. Subsequently, we develop a technique to design a constrained analog precoder that mimics the obtained unconstrained analog precoder under phase shifter constraints. The main idea is to adopt an additional baseband precoding matrix, which we call a compensation matrix. We analyze the SLNR loss due to the proposed hybrid precoding compared to fully digital precoding, and determine which factors have a significant impact on this loss. In the simulations, we show that if the channel is spatially correlated and the number of users is smaller than the number of RF chains, the SLNR loss becomes negligible compared to fully digital precoding. The main benefit of our method stems from the use of spatial channel matrices in such a way that not only is each user's desired signal considered, but also the inter-user interference is incorporated in the analog precoder design.**

***Index Terms*—Hybrid precoding, multi-user MIMO, massive MIMO, covariance matrix, cellular communication, antenna arrays.**

I. Introduction

MASSIVE MIMO systems increase cellular spectral efficiency by employing many antennas at the base station to support multiuser MIMO on the uplink and downlink. Such systems require a higher number of radio-frequency (RF) chains resulting in high mixed signal complexity and larger power consumption at the base-station. Hybrid analog/digital transmit precoding aims to alleviate these issues by using a reduced set of RF chains mapped via analog beamforming to a large number of physical antennas. The hybrid precoding method essentially divides the precoding process at the transmitter between the analog RF and digital baseband part. This technique was first investigated for general MIMO systems in [1], [2] and was later applied to millimeter wave systems in [3]–[6]. Hybrid precoding originally focused on beamforming and single-user MIMO techniques, but it also extends to multiuser MIMO configurations [7]–[16].

Most prior work on hybrid precoding for massive MIMO systems or multiuser millimeter wave systems assumes full channel state information at transmitter (CSIT) for all antennas when designing the analog precoder [7]–[13]. Such an approach has two fundamental limitations: First, it is inherently difficult to obtain reliable instantaneous CSIT estimates in hybrid systems. Even in time-division duplexing (TDD) systems where channel reciprocity can be exploited, the hybrid structure makes it difficult to estimate the *entire* channel matrix for all antennas since the estimator in the baseband can only see a low(er)-dimensional pre-combined channel through few(er) RF chains. To overcome this challenge, some clever channel estimation techniques for the hybrid structure were proposed by using compressive sensing methods in the spatially sparse channel [17], [18]. These techniques, however, require more measurements than a fully digital structure and require sparsity in the channel. Consequently, these methods work well only for time-invariant channels or very slowly varying channels. Second, it is hard to directly apply this approach to wideband OFDM systems where the channel is frequency selective. While the digital baseband precoder can be adapted to the channel of each subcarrier in the frequency domain, the same analog RF precoder must be applied on all the subcarriers as the analog RF precoder operates in the time domain [19]. This is a distinguishing constraint on wideband systems compared to narrowband systems. Since prior work [7]–[13] that assumes full CSIT in narrowband systems focuses on the joint optimization between the baseband precoder and the analog precoder for each channel realization, this approach will produce a different analog precoder for each subcarrier if applied to frequency selective wideband channels. This fails to satisfy the constraint on the analog precoder for wideband systems, and thus the approach in [7]–[13] cannot be directly applied to wideband MIMO-OFDM systems.

A reasonable alternative to instantaneous CSIT is to use long-term channel statistics, in particular the spatial channel covariance, to configure the analog precoder. Firstly, spatial channel covariances vary over a longer time scale compared to the instantaneous channels, which makes it easier to estimate [20]. Secondly, the spatial channel covariance is uniform across all subcarriers, providing a good match to the problem of designing

Manuscript received December 10, 2016; revised March 31, 2017; accepted April 23, 2017. Date of publication May 4, 2017; date of current version May 22, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Eleftherios Kofidis. *(Corresponding author: Robert W. Heath, Jr.)*

S. Park, J. Park, and R. W. Heath, Jr., are with the Wireless Networking and Communication Group, Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78701 USA (e-mail: swpark96@utexas.edu; jeonghun@utexas.edu; rheath@utexas.edu).

A. Yazdan is with the Connectivity Lab., Facebook, Menlo Park, CA 94025 USA (e-mail: ayp@fb.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2017.2701321

one analog precoder for all subcarriers [21]. Motivated by this, we propose a hybrid multiuser precoding algorithm where the analog precoder is designed using each user's spatial channel covariance of its entire channel matrix.

Limited work has been done on the multiuser hybrid precoding technique using spatial channel covariance matrices in the analog precoder design. In [15], users are divided into groups based on their locations such that the users in each group have an identical spatial channel covariance matrix. After grouping, a so-called two-stage precoding is performed: a prebeamforming matrix that eliminates inter-group interference by using spatial channel covariance matrices, and a multiuser MIMO precoding matrix that removes intra-group interference by using per-user low-dimensional channel matrices. This two-stage approach, however, does not explicitly take the hybrid analog/digital architecture into consideration, thereby making it difficult in general to apply the prebeamfomer to the analog RF part that consists of phase shifters. Another problem of this approach is its assumption that all users in a group have the same covariance matrix, which requires a very large number of users. In addition, the columns (corresponding to the RF chains in the hybrid structure) in the prebeamforming matrix are evenly assigned to groups. In [22], the hybrid precoding is developed under the phase shifter constraint, and the user-grouping concept in [15] is extended to a more general case where different numbers of RF chains are dynamically assigned to different groups. While the hybrid precoding techniques in [15], [22] are limited to the assumption that there exist groups in which users share the same covariance matrices, a general hybrid precoding technique is proposed without user grouping in [16], where each RF chain is dedicated to a user. In this case, each analog precoding vector associated with each RF chain is constructed from each user's single dominant eigenvector of the covariance matrix. A similar approach is found in the hybrid precoding techniques that use full CSIT in [8], [10], [11], [23]. This approach, however, does not consider interference in the design of analog beamforming, only maximizing each user's desired power. In addition, the number of assigned users must be the same as the number of RF chains.

In this paper, we propose a new hybrid precoding algorithm for multiuser massive MIMO systems. In our work, we use regularized zero-forcing (RZF) precoding [24] as the digital baseband precoder. Although RZF precoding is not optimal in a capacity-maximizing sense, it is an attractive alternative to dirty paper coding (DPC) [25], thanks to its low complexity. Since the RZF, which is also known as transmit Wiener filter or transmit MMSE precoding (see [26] and references therein), is optimal in the sense of maximizing the signal-to-leakage-plus-noise ratio (SLNR) [26]–[28], the SLNR is a reasonable metric for analysis or precoder design and has been frequently used in the work on massive MIMO systems. For example, bounds on the ergodic rate when using RZF in massive MIMO systems are analyzed by using the SLNR in [29], and the optimal user loading to maximize the sum rate in massive MIMO systems is derived based on the SLNR in [30]. In addition, the work in [31], [32] proposes precoding techniques based on the SLNR in multiuser MIMO systems [31] or massive MIMO systems [32]. Therefore, we adopt the SLNR maximization to design the hybrid precoders. Unlike conventional RZF precoders that solely operate in the baseband with full CSIT, our hybrid precoding tries to maximize the SLNR with a combination of an analog and a digital precoder. In addition, our design of the analog precoder only requires spatial channel covariance matrices of users instead of full CSIT.

Our design process is as follows. We first develop an unconstrained analog precoder without the constraint that analog precoding is realized with phase shifters. The unconstrained analog precoder is designed to aid the baseband precoder to maximize the SLNR, only with the knowledge of users' spatial channel covariance matrices. Subsequently, taking the phase shifter constraint into account, we divide the obtained unconstrained analog precoding matrix into two separate matrices: a constrained analog precoding matrix and an additional baseband precoding matrix, which we call a compensation matrix. The compensation matrix depends only on the constrained analog precoder, and its role is to mitigate the effect caused by using phase shifters. Using the compensation matrix, the constrained analog precoder is optimized so that the combination of the constrained analog precoder and the compensation matrix is as similar to the unconstrained analog precoder as possible. Since the compensation matrix is determined by the constrained analog precoder, it does not require instantaneous CSIT although this compensation matrix operates in the baseband.

A distinguishing feature from the prior work in [16] is that our work can be applied to the case when the number of assigned users is less than that of the RF chains, without dedicating an RF chain to a user (or a group of RF chains to a user group [15]). Dedicating each RF chain to a specific user is not efficient in that its applicability is restricted to the case when the number of users is equal to the number of RF chains. In typical cellular systems, the number of assigned users tends to vary over time while the number of RF chains is fixed. This makes our proposed technique more beneficial in realistic scenarios. Similar work has been recently done in [33] where the proposed scheme exploits only statistical channel state information without dedicating an RF chain to a user nor a group of RF chains to a user group, similar to our work. Although the finding in [33] also accords closely with ours, there are some differences. While the work in [33] considers codebook-based hybrid precoding and focuses only on designing the analog precoder, our work does not confine its application to codebook-based hybrid precoding and deals with the design of the baseband precoder as well as the analog precoder.

We analyze the SLNR loss caused by the hybrid precoding compared to the fully digital RZF precoding in massive MIMO systems for various channel conditions. In single-user MIMO systems, the loss from hybrid precoding is negligible if the number of channel paths is smaller than the number of RF chains in a spatially sparse channel. The loss in multiuser MIMO systems depends, however, not only on the number of channel paths for each user but also on the number of users. Therefore, hybrid precoding in multiuser MIMO is likely to have higher loss than in the single-user MIMO. Our finding, however, reveals that the loss from hybrid precoding is still low in spatially correlated

channels if the number of users is small enough compared to the number of RF chains.

The proposed hybrid precoding designs are evaluated by simulations in terms of sum spectral efficiency. The results show that the developed hybrid precoding design outperforms prior work that dedicates a RF chain to a user. Simulation results also illustrate that the proposed constrained analog precoder with the compensation matrix results in almost the same spectral efficiency as the unconstrained analog precoder. In addition, both simulation and analysis results indicate that the loss caused by the hybrid architecture can be low even though the proposed analog precoder design does not require instantaneous CSIT. This promotes the employment of massive MIMO systems in practical cellular networks.

The rest of the paper is organized as follows: In Section II we introduce the system and channel models. In Section III, we obtain the unconstrained analog precoder to maximize the SLNR by using the spatial channel covariance matrices. In what follows, we propose a technique to mimic the unconstrained analog precoder in Section IV. In Section V, we analyze the SLNR loss in the proposed strategy. In Section VI, simulation results validate the proposed method and Section VII concludes the paper.

Notation: We use the following notation throughout this paper: $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, a is a scalar, and $\mathbb{A}$ is a set. $|a|$ and $\measuredangle a$ are the magnitude and phase of the complex number a. $\|\mathbf{A}\|_{\mathrm{F}}$ is its Frobenius norm, and $\mathbf{A}^T$, $\mathbf{A}^*$, and $\mathbf{A}^{-1}$ are its transpose, Hermitian (conjugate transpose), and inverse, respectively. $[\mathbf{A}]_{m,n}$ is the (m,n)-th element of the matrix $\mathbf{A}$. $\measuredangle\mathbf{A}$ is a matrix whose (m,n)-th element equals $e^{j[\mathbf{A}]_{m,n}}$. $\mathbf{I}$ is an identity matrix and $\mathbf{0}$ is a matrix whose elements are all zeros. Subscripts will be used to denote their dimensions such as $\mathbf{I}_N$ and $\mathbf{0}_{N\times M}$, if necessary. $\mathcal{CN}(\mathbf{m},\mathbf{R})$ is a complex Gaussian random vector with mean $\mathbf{m}$ and covariance $\mathbf{R}$. $\mathbb{E}[\cdot]$ is used to denote expectation.

II. System and Channel Models

Consider the downlink of a massive MIMO system where a base station (BS) equipped with N antennas and $M(\leq N)$ RF chains communicates with $U(\leq M)$ users with a single antenna. Let $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N\times M}$, $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{M\times U}$, $\mathbf{P} \in \mathbb{R}^{U\times U}$, and $\mathbf{s} \in \mathbb{C}^{U\times 1}$ be the analog RF precoder, a digital baseband precoder, a power control matrix, and a signal vector. The received signal is given by

$$\mathbf{y} = \mathbf{H}^*\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{P}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \in \mathbb{C}^{U\times 1} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ is circularly symmetric complex Gaussian noise, $\mathbf{P}$ is a diagonal matrix to maintain the total transmit power P_{tx}, and $\mathbf{H}^* = \left[\mathbf{h}_1 \cdots \mathbf{h}_U\right]^* \in \mathbb{C}^{U\times N}$ is the aggregate downlink channel matrix composed of each user's channel vector $\mathbf{h}_u, \forall u$.

We consider three constraints on hybrid precoding in our system model.

- Constraint 1: The number of RF chains is less than the number of antennas.
- Constraint 2: $\mathbf{F}_{\mathrm{RF}}$ only depends on users' spatial channel covariance matrices $\mathbf{R}_u = \mathbb{E}\left[\mathbf{h}_u\mathbf{h}_u^*\right], \forall u$.
- Constraint 3: $\mathbf{F}_{\mathrm{RF}}$ is composed of phase shifters, i.e., all the elements in $\mathbf{F}_{\mathrm{RF}}$ have the same amplitudes.

We assume that $\mathbf{R}_u$ is known to the BS through covariance estimation for the hybrid structure, see e.g. [34]–[36]. Note that $\mathbf{F}_{\mathrm{RF}}$ is designed by using only $\mathbf{R}_u$'s, not $\mathbf{h}_u$'s as shown in Constraint 2. This is different from prior work [7]–[13] where $\mathbf{F}_{\mathrm{RF}}$ depends on the instantaneous channel $\mathbf{h}_u$'s.

We focus on the ideal covariance estimation case where the ideal $\mathbf{R}_u$ is known to the BS without estimation error. The impact of the covariance estimation error on the performance of our hybrid precoding design is out of the scope of this paper and remains as a future research topic. Since covariance estimation error will affect spectral efficiency differently from channel estimation error, this will be an interesting topic for future work.

In regard to modeling each user's channel vector $\mathbf{h}_u$, we use a Kronecker correlation model for algorithm design due to its analytical tractability in Section III. The channel $\mathbf{h}_u$ is modeled as $\mathbf{h}_u = \mathbf{R}_u^{\frac{1}{2}}\mathbf{h}_{w,u}$ where $\mathbf{h}_{w,u}$ has identically independent distributed (IID) complex entries of zero mean and unit variance. The $\mathbf{R}_u$'s are regarded as deterministic variables while $\mathbf{h}_u$ and $\mathbf{n}$ as random variables in Section III. For numerical evaluations in Section VI, we adopt a simple geometry-based channel model to investigate the relationship between the spatial channel sparsity and the performance loss caused by the hybrid structure. The details of the associated simulation parameters will be described in Section VI.

III. Hybrid Precoding Using Spatial Channel Covariance Matrices

In this section, we design hybrid precoder under Constraint 1 and 2; adding Constraint 3 will be considered in Section IV.

For intuition, we first introduce the prior design technique in [16] where each RF chain is dedicated to a user. In the design of an analog precoder, statistical eigen beamforming is used where

$$\mathbf{F}_{\mathrm{RF}} = \left[\mathbf{v}_{1,\max} \cdots \mathbf{v}_{U,\max}\right], \tag{2}$$

and $\mathbf{v}_{u,\max}$ is the dominant eigenvector of $\mathbf{R}_u$. Once $\mathbf{F}_{\mathrm{RF}}$ is designed, the baseband precoder performs conventional multiuser MIMO techniques such as zero-forcing (ZF) or regularized zero-forcing (RZF) with respect to the combined effective channel $\mathbf{H}_{\mathrm{eff}}^* = \mathbf{H}^*\mathbf{F}_{\mathrm{RF}}$. This approach is similar to [8], [10], [11], [23] using full CSIT. The rationale behind this technique is to maximize the long-term average power of the desired signal in the analog part. The main drawback of this approach is that interference is neglected, which results in performance degradation unless the channel is ideally orthogonal. Moreover, this technique cannot be directly applied when $U < M$.

Now we explain our design. We will focus on the RZF case throughout the paper where

$$\begin{aligned}\mathbf{F}_{\mathrm{BB}} &= \left[\mathbf{f}_{\mathrm{BB},1} \cdots \mathbf{f}_{\mathrm{BB},U}\right] \\ &= \left(\mathbf{F}_{\mathrm{RF}}^*\mathbf{H}\mathbf{H}^*\mathbf{F}_{\mathrm{RF}} + \beta\mathbf{I}\right)^{-1}\mathbf{F}_{\mathrm{RF}}^*\mathbf{H},\end{aligned} \tag{3}$$

and $\beta = \frac{U}{\rho}$ is a regularization parameter [24], where $\rho = \frac{P_{\mathrm{tx}}}{\sigma^2}$ denotes the transmit SNR. We consider an equal power strategy that makes each user's power equal after precoding including both $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$, so the u-th diagonal element of $\mathbf{P}$ in (1) is

$$p_u = \frac{\sqrt{P_{\mathrm{tx}}}}{\sqrt{U}\|\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{bb},u}\|} = \sqrt{\frac{P_{\mathrm{tx}}}{U\mathbf{h}_u^*\mathbf{W}^2\mathbf{h}_u}}, \tag{4}$$

where $\mathbf{W} = \mathbf{F}_{\mathrm{RF}}(\mathbf{F}_{\mathrm{RF}}^*\mathbf{H}\mathbf{H}^*\mathbf{F}_{\mathrm{RF}} + \frac{U}{\rho}\mathbf{I})^{-1}\mathbf{F}_{\mathrm{RF}}^*$. The instantaneous SLNR is given by

$$\begin{aligned}\mathrm{SLNR}_u &= \frac{p_u^2|\mathbf{h}_u^*\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{bb},u}|^2}{\sum_{i\neq u}p_u^2|\mathbf{h}_i^*\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{bb},u}|^2+\sigma^2}\\ &= \frac{\mathbf{h}_u^*\mathbf{W}\mathbf{h}_u\mathbf{h}_u^*\mathbf{W}\mathbf{h}_u}{\sum_{i\neq u}\mathbf{h}_u^*\mathbf{W}\mathbf{h}_i\mathbf{h}_i^*\mathbf{W}\mathbf{h}_u+\frac{U\sigma^2}{P_{\mathrm{tx}}}\mathbf{h}_u^*\mathbf{W}^2\mathbf{h}_u}\\ &= \frac{\mathbf{h}_u^*\mathbf{W}\mathbf{h}_u\mathbf{h}_u^*\mathbf{W}\mathbf{h}_u}{\mathbf{h}_u^*\mathbf{W}\left(\mathbf{H}\mathbf{H}^*-\mathbf{h}_u\mathbf{h}_u^*+\frac{U}{\rho}\mathbf{I}\right)\mathbf{W}\mathbf{h}_u}.\end{aligned} \tag{5}$$

The goal is to find $\mathbf{F}_{\mathrm{RF}}$ to maximize the SLNR in (5). Instead of assigning each column of $\mathbf{F}_{\mathrm{RF}}$ to each user as in the prior work, our design configures $\mathbf{F}_{\mathrm{RF}}$ from an orthonormal basis spanning a subspace that maximizes the SLNR. Therefore, there is no constraint such as $U = M$, so this approach can be applied for the case of $U < M$ as well. Later, we will show that allocating fewer than M users is better than allocating M users when M RF chains are given. Let $\mathbf{F}_{\mathrm{RF}}$ be a rank-$\tilde{M}$ matrix where $\tilde{M} \leq M$. Then, $\mathbf{F}_{\mathrm{RF}}$ can be represented as $\mathbf{F}_{\mathrm{RF}} = \mathbf{V}\mathbf{A}$ where $\mathbf{V} \in \mathbb{C}^{N\times\tilde{M}}$ is a semi-unitary matrix such that $\mathbf{V}^*\mathbf{V} = \mathbf{I}$ and $\mathbf{A} \in \mathbb{C}^{\tilde{M}\times M}$ is a rank-$\tilde{M}$ matrix. This can be easily proved by applying SVD. In the following proposition, we show that the SLNR in (5) has a maximum value when $\mathbf{A}$ is semi-unitary, i.e., $\mathbf{F}_{\mathrm{RF}}$ needs to be a multiplication of two semi-unitary matrices such that $\mathbf{F}_{\mathrm{RF}} = \mathbf{V}\mathbf{U}^*$ where $\mathbf{V}^*\mathbf{V} = \mathbf{U}^*\mathbf{U} = \mathbf{I}$ to maximize the SLNR.

Proposition 1: Suppose that $\mathbf{F}_{\mathrm{RF}}$ is decomposed as $\mathbf{F}_{\mathrm{RF}} = \mathbf{V}\mathbf{A}$ where $\mathbf{V} \in \mathbb{C}^{N\times\tilde{M}}$ is a semi-unitary matrix such that $\mathbf{V}^*\mathbf{V} = \mathbf{I}$ and $\mathbf{A} \in \mathbb{C}^{\tilde{M}\times M}$ is a rank-$\tilde{M}$ matrix. If $\mathbf{V}$ and P_{tx} is given, the SLNR in (5) is maximized when $\mathbf{A}$ is semi-unitary, i.e., $\mathbf{A}\mathbf{A}^* = \mathbf{I}$, and the maximum value becomes

$$\mathrm{SLNR}_u = \mathbf{h}_u^*\mathbf{V}\left(\sum_{i\neq u}^{U}\mathbf{V}^*\mathbf{h}_i\mathbf{h}_i^*\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\mathbf{V}^*\mathbf{h}_u, \tag{6}$$

for any semi-unitary matrix $\mathbf{A}$.

Proof: See Appendix A. ■

Since the SLNR is independent of $\mathbf{A}$ as long as $\mathbf{A}$ is semi-unitary, let us focus on constructing $\mathbf{V}$ to maximize the SLNR. Note that the SLNR in (6) is a random variable due to $\mathbf{h}_i$. The random variable SLNR, however, converges to a deterministic value as the number of antennas becomes large. Let $\mathbf{h}_i = \mathbf{R}_i^{\frac{1}{2}}\mathbf{h}_{w,i} = \sqrt{N}\mathbf{R}_i^{\frac{1}{2}}\mathbf{g}_i$, where $\mathbf{g}_i$ has IID complex entries with zero mean and variance of $1/N$. Then, as N goes to infinity, the SLNR in (6) converges to

$$\begin{aligned}&\mathrm{SLNR}_u\\ &= N\mathbf{g}_u^*\mathbf{R}_u^{\frac{1}{2}}\mathbf{V}\left(N\sum_{i\neq u}^{U}\mathbf{V}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{g}_i\mathbf{g}_i^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\mathbf{V}^*\mathbf{R}_u^{\frac{1}{2}}\mathbf{g}_u\\ &\xrightarrow{a.s.}\mathrm{Tr}\left(\mathbf{R}_u^{\frac{1}{2}}\mathbf{V}\left(N\sum_{i\neq u}^{U}\mathbf{V}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{g}_i\mathbf{g}_i^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\mathbf{V}^*\mathbf{R}_u^{\frac{1}{2}}\right)\\ &\xrightarrow{a.s.}\mathrm{Tr}\left(\mathbf{R}_u^{\frac{1}{2}}\mathbf{V}\left(N\sum_{i=1}^{U}\mathbf{V}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{g}_i\mathbf{g}_i^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\mathbf{V}^*\mathbf{R}_u^{\frac{1}{2}}\right)\\ &= \mathrm{Tr}\left(\mathbf{V}^*\mathbf{R}_u\mathbf{V}\left(N\sum_{i=1}^{U}\mathbf{V}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{g}_i\mathbf{g}_i^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right),\end{aligned} \tag{7}$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence as $N \to \infty$ [37]. In (7), the first convergence comes from the trace lemma [38], and the second convergence comes from the rank-1 perturbation lemma [38].

By [37, Th. 1], the random variable SLNR converges to a deterministic SLNR value

$$\mathrm{SLNR}_u \xrightarrow{a.s.} \gamma_u, \tag{8}$$

where $\gamma_1, \ldots, \gamma_U$ are the unique nonnegative solution of

$$\gamma_u = \mathrm{Tr}\left(\mathbf{V}^*\mathbf{R}_u\mathbf{V}\left(\sum_{j=1}^{U}\frac{\mathbf{V}^*\mathbf{R}_j\mathbf{V}}{1+\gamma_j}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right). \tag{9}$$

The solution of $\gamma_1, \ldots, \gamma_U$ can be obtained fixed point equations [37], [38] as $\gamma_u = \lim_{t\to\infty}\gamma_u^{(t)}$ where

$$\gamma_u^{(t)} = \mathrm{Tr}\left(\mathbf{V}^*\mathbf{R}_u\mathbf{V}\left(\sum_{j=1}^{U}\frac{\mathbf{V}^*\mathbf{R}_j\mathbf{V}}{1+\gamma_j^{(t-1)}}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right). \tag{10}$$

Let us consider the optimization problem that maximizes the asymptotic SLNR averaged over all users as

$$\begin{aligned}&\max_{\mathbf{V}\in\mathbb{V}}\frac{1}{U}\sum_{u=1}^{U}\gamma_u\\ &\text{s.t.}\quad \gamma_u = \mathrm{Tr}\left(\mathbf{V}^*\mathbf{R}_u\mathbf{V}\left(\sum_{j=1}^{U}\frac{\mathbf{V}^*\mathbf{R}_j\mathbf{V}}{1+\gamma_j}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right), \forall u,\end{aligned} \tag{11}$$

where $\mathbb{V} = \{\mathbf{X} \mid \mathbf{X}^*\mathbf{X} = \mathbf{I}, \mathbf{X} \in \mathbb{C}^{N\times m}, m = 1, \ldots, M\}$. This is difficult to solve directly due to the U fixed point equations in (11). Instead, we resort to a simplified problem where we assume that all users have the same SLNR as $\gamma_1 = \cdots = \gamma_U$

$= \gamma = \frac{1}{U}\sum_{u=1}^{U}\gamma_u$. Then, the optimization problem becomes

$$\max_{\mathbf{V}\in\mathbb{V}} \gamma$$
$$\text{s.t.} \quad \gamma = \mathrm{Tr}\left(\mathbf{V}^*\mathbf{R}_{\mathrm{tot}}\mathbf{V}\left(\frac{U\mathbf{V}^*\mathbf{R}_{\mathrm{tot}}\mathbf{V}}{1+\gamma}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right), \quad (12)$$

where $\mathbf{R}_{\mathrm{tot}} = \frac{1}{U}\sum_{u=1}^{U}\mathbf{R}_u$. Let $\mathbf{V}^*\mathbf{R}_{\mathrm{tot}}\mathbf{V}$ be decomposed as $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$ by eigenvalue decomposition and have eigenvalues of $\nu_1,\ldots,\nu_M$ in descending order. Then, γ is rewritten as

$$\gamma = \frac{1}{U}\mathrm{Tr}\left(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^*\left(\frac{\mathbf{U}\mathbf{\Lambda}\mathbf{U}^*}{1+\gamma}+\frac{1}{\rho}\mathbf{I}\right)^{-1}\right)$$
$$= \frac{1}{U}\sum_{m=1}^{\tilde{M}}\frac{1}{\frac{1}{1+\gamma}+\frac{1}{\rho\nu_m}} \quad (13)$$

where $\tilde{M} = \min(M, \mathrm{rank}(\mathbf{R}_{\mathrm{tot}}))$. Then, the solution to (12) is given in the following proposition.

Proposition 2: The $\mathbf{V}$ that solves the maximization in (12) is the matrix whose columns are composed of $\tilde{M}$ eigenvectors associated with the $\tilde{M}$ largest eigenvalues of $\mathbf{R}_{\mathrm{tot}} = \frac{1}{U}\sum_{u=1}^{U}\mathbf{R}_u$ where $\tilde{M} = \min(M, \mathrm{rank}(\mathbf{R}_{\mathrm{tot}}))$.

Proof: See Appendix B. ■

Proposition 2 indicates that the analog precoding $\mathbf{F}_{\mathrm{RF}}$ that results from the optimization problem in (12) uses the $\tilde{M}$ dominant eigenvectors of $\mathbf{R}_{\mathrm{tot}} = \frac{1}{U}\sum_{u=1}^{U}\mathbf{R}_u$, i.e., the sum of the spatial covariance matrices of U users.

Although the derived solution is based on a simplified optimization problem assuming that each user's SLNR asymptotically converges to the average value over users in the large antenna array regime, we will show in Section VI that this solution outperforms the prior work using (2) and has spectral efficiency close to that of the fully digital precoding in spatially correlated channels. In addition, we will show that the proposed solution has exactly the same spectral efficiency as that of the fully digital precoding if $\mathbf{R}_{\mathrm{tot}}$ is rank-deficient and its rank is less than or equal to M. This will be discussed further in Section V.

Even in the finite antenna regime where the SLNR does not converge, the proposed analog precoder is beneficial in the sense that the proposed technique maximizes a lower bound of the expectation of the SLNR averaged over U users. The expectation of the average SLNR over U users is expressed in (14) as shown at the bottom of this page, where (a) comes from the fact that $\mathbf{h}_{w,u}$'s are independent, (b) comes from the fact that $\mathbb{E}[\mathbf{A}^{-1}] - (\mathbb{E}[\mathbf{A}])^{-1}$ is positive semidefinite for a positive semidefinite matrix $\mathbf{A}$ and $\mathrm{Tr}(\mathbf{AB}) \geq 0$ for positive matrices $\mathbf{A}$ and $\mathbf{B}$, (c) comes from the definition of $\mathbf{R}_u$, and (d) comes from the fact that $\mathbf{A}^{-1} - (\mathbf{A}+\mathbf{B})^{-1}$ is positive semidefinite for positive semidefinite matrices $\mathbf{A}$ and $\mathbf{B}$. With the same notation used in (12) and (13), the lower bound in (14) can be expressed as

$$\mathbb{E}\left[\frac{1}{U}\sum_{u=1}^{U}\mathrm{SLNR}_u\right]$$
$$\geq \mathrm{Tr}\left(\frac{1}{U}\mathbf{V}^*\mathbf{R}_{\mathrm{tot}}\mathbf{V}\left(\mathbf{V}^*\mathbf{R}_{\mathrm{tot}}\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right)$$
$$= \sum_{m=1}^{\tilde{M}}\frac{1}{U+\frac{\rho}{\nu_m}}. \quad (15)$$

This lower bound in (15) has a similar form to (13). As a result, the $\mathbf{V}$ that maximizes the expected average SLNR is the same as the optimal solution in Proposition 2.

The proposed approach has some advantages over the prior work. While each RF chain is dedicated to one user in the analog precoding in the prior work in [8], [10], [11], [16], [23], in our case all the RF chains construct a subspace for all users as a whole in the proposed analog precoding. For this reason, there is no limitation on assigning the exactly same number of the users to the number of the RF chains, providing a wider range of applicability of the proposed method.

$$\mathbb{E}\left[\frac{1}{U}\sum_{u=1}^{U}\mathrm{SLNR}_u\right] = \mathrm{Tr}\left(\frac{1}{U}\sum_{u=1}^{U}\mathbb{E}\left[\mathbf{V}^*\mathbf{R}_u^{\frac{1}{2}}\mathbf{h}_{w,u}\mathbf{h}_{w,u}^*\mathbf{R}_u^{\frac{1}{2}}\mathbf{V}\left(\sum_{i\neq u}^{U}\mathbf{V}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{h}_{w,i}\mathbf{h}_{w,i}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right]\right)$$
$$\overset{(a)}{=} \mathrm{Tr}\left(\frac{1}{U}\sum_{u=1}^{U}\mathbb{E}\left[\mathbf{V}^*\mathbf{R}_u^{\frac{1}{2}}\mathbf{h}_{w,u}\mathbf{h}_{w,u}^*\mathbf{R}_u^{\frac{1}{2}}\mathbf{V}\right]\mathbb{E}\left[\left(\sum_{i\neq u}^{U}\mathbf{V}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{h}_{w,i}\mathbf{h}_{w,i}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right]\right)$$
$$\overset{(b)}{\geq} \mathrm{Tr}\left(\frac{1}{U}\sum_{u=1}^{U}\mathbb{E}\left[\mathbf{V}^*\mathbf{R}_u^{\frac{1}{2}}\mathbf{h}_{w,u}\mathbf{h}_{w,u}^*\mathbf{R}_u^{\frac{1}{2}}\mathbf{V}\right]\left(\mathbb{E}\left[\sum_{i\neq u}^{U}\mathbf{V}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{h}_{w,i}\mathbf{h}_{w,i}^*\mathbf{R}_i^{\frac{1}{2}}\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right]\right)^{-1}\right)$$
$$\overset{(c)}{=} \mathrm{Tr}\left(\frac{1}{U}\sum_{u=1}^{U}\mathbf{V}^*\mathbf{R}_u\mathbf{V}\left(\sum_{i\neq u}^{U}\mathbf{V}^*\mathbf{R}_i\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right)$$
$$\overset{(d)}{\geq} \mathrm{Tr}\left(\frac{1}{U}\sum_{u=1}^{U}\mathbf{V}^*\mathbf{R}_u\mathbf{V}\left(\sum_{i=1}^{U}\mathbf{V}^*\mathbf{R}_i\mathbf{V}+\frac{U}{\rho}\mathbf{I}\right)^{-1}\right), \quad (14)$$

Note that the proposed technique can be applied without modification even when there are some users sharing the same covariance matrix. Suppose that all users have the same covariance matrix. This is an extreme case where there exists only one user group sharing the covariance matrix. The role of the analog precoder is to concentrate the transmit power on the local scatterers shared by this user group. Even in this extreme case, the RZF baseband precoder can still mitigate inter-user interference if the number of users is less than the number of dominant eigenvalues of the covariance matrix, which is closely related to the number of dominant local scatterers. In addition, note that the proposed hybrid precoding has the same performance as the fully digital precoding if the rank of the covariance matrix is less than or equal to M.

Fairness among users is not considered in this paper. There are two different approaches to take fairness into consideration for conventional multiuser MIMO. One is to design precoders based on some fairness criteria, e.g., max-min-rate criterion or max-sum-rate criterion under some quality-of-service (QoS) constraints such as a minimum user rate requirement. The other is to combine unfair precoding with fair scheduling, e.g., the combination of a sum-rate-maximizing precoder and a proportional fair scheduling. Either precoder design considering fairness or joint optimization of scheduling and precoder under fairness constraints will be an interesting topic for future work.

IV. Hybrid Precoding Under Phase Shifter Constraint

In this section, we add Constraint 3 into our precoder design. Specifically, we propose a technique to mimic $\mathbf{F}_{\mathrm{RF}}$ obtained in Section III under the phase shifter constraint. We will refer to the unconstrained $\mathbf{F}_{\mathrm{RF}}$ derived in Section III as $\mathbf{F}_{\mathrm{RF,UC}}$ and its constrained version as $\mathbf{F}_{\mathrm{RF,C}}$.

In prior work [8], the constrained precoder was found by solving

$$\min_{\mathbf{F}_{\mathrm{RF,C}},|[\mathbf{F}_{\mathrm{RF,C}}]_{i,j}|=\frac{1}{\sqrt{N}}} \left\|\mathbf{F}_{\mathrm{RF,UC}}-\mathbf{F}_{\mathrm{RF,C}}\right\|_{\mathrm{F}}^{2}. \tag{16}$$

The precoder $\mathbf{F}_{\mathrm{RF,C}}$ that minimizes the Frobenius norm of the difference between $\mathbf{F}_{\mathrm{RF,C}}$ and $\mathbf{F}_{\mathrm{RF,UC}}$ is a reasonable approximation of $\mathbf{F}_{\mathrm{RF,UC}}$ [3]. The solution of (16) is given by $[\mathbf{F}_{\mathrm{RF,C}}^{(\mathrm{opt})}]_{i,j}=\frac{1}{\sqrt{N}}e^{j\angle([\mathbf{F}_{\mathrm{RF,UC}}]_{i,j})}$. The weakness of this approach is that $\mathbf{F}_{\mathrm{RF,C}}$ loses the orthogonality that $\mathbf{F}_{\mathrm{RF,UC}}$ retains, thereby leading to performance degradation. In our design, $\mathbf{F}_{\mathrm{RF,UC}}$ needs to have a multiplicative form of two semi-unitary matrices to maximize the SLNR according to Proposition 1.

To overcome this weakness, we apply a compensation matrix in the baseband part to restore the orthogonality lost in the analog part as shown in Fig. 2. (We note that Fig. 1 indicates the hybrid precoding using spatial channel covariance in the analog precoder part without Constraint 3.) Let $\mathbf{F}_{\mathrm{RF,C}}=\mathbf{U}_{\mathrm{RF,C}}\mathbf{D}_{\mathrm{RF,C}}\mathbf{V}_{\mathrm{RF,C}}^{*}$ by compact SVD such that $\mathbf{D}_{\mathrm{RF,C}}$ is an $r\times r$ diagonal matrix where $r=\mathrm{rank}(\mathbf{F}_{\mathrm{RF,C}})$. Then, the compensation matrix $\mathbf{F}_{\mathrm{CM}}$ is given by

$$\mathbf{F}_{\mathrm{CM}}=\mathbf{V}_{\mathrm{RF,C}}\mathbf{D}_{\mathrm{RF,C}}^{-1}\mathbf{V}_{\mathrm{RF,C}}^{*}. \tag{17}$$

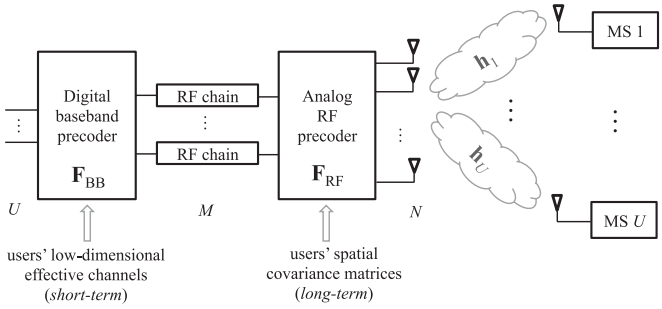
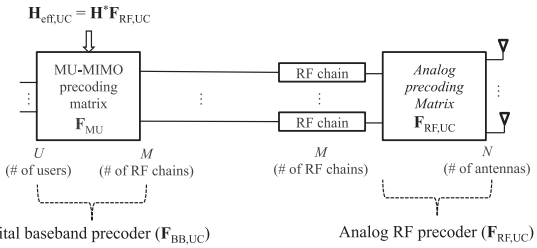
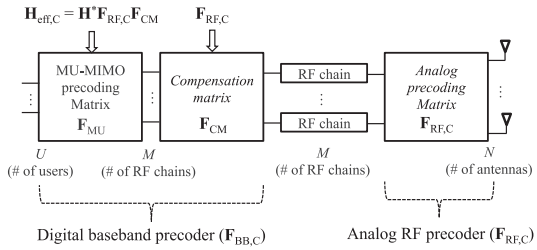


Fig. 1. Hybrid precoding for massive MIMO systems using spatial channel covariance matrices in the analog precoder design.



(a) Hybrid architecture for massive MIMO systems without phase shifter constraints (composed of an unconstrained analog precoding matrix and a RZF baseband precoding matrix)



(b) Hybrid architecture for massive MIMO systems with phase shifter constraints (composed of a constrained analog precoding matrix, a compensation matrix, and a RZF baseband precoding matrix)

Fig. 2. Proposed hybrid precoding structure.

The combination of $\mathbf{F}_{\mathrm{RF,C}}$ and $\mathbf{F}_{\mathrm{CM}}$ becomes a multiplicative form of two semi-unitary matrices as

$$\begin{aligned}\mathbf{F}_{\mathrm{RF,C}}\mathbf{F}_{\mathrm{CM}}&=\mathbf{U}_{\mathrm{RF,C}}\mathbf{D}_{\mathrm{RF,C}}\mathbf{V}_{\mathrm{RF,C}}^{*}\mathbf{V}_{\mathrm{RF,C}}\mathbf{D}_{\mathrm{RF,C}}^{-1}\mathbf{V}_{\mathrm{RF,C}}^{*}\\&=\mathbf{U}_{\mathrm{RF,C}}\mathbf{V}_{\mathrm{RF,C}}^{*}.\end{aligned} \tag{18}$$

Applying this compensation matrix allows further improvement in designing $\mathbf{F}_{\mathrm{RF,C}}$ by using the following property.

Proposition 3: Let $\mathbf{F}_{\mathrm{RF,UC}}=\mathbf{V}\mathbf{U}^{*}\in\mathbb{C}^{N\times M}$ denote an optimal unconstrained analog precoding matrix with rank $\tilde{M}$ where $\mathbf{V}^{*}\mathbf{V}=\mathbf{U}^{*}\mathbf{U}=\mathbf{I}_{\tilde{M}}$. Suppose that $\mathbf{F}_{\mathrm{RF,UC}}\mathbf{A}$ is used in place of the constrained analog precoder $\mathbf{F}_{\mathrm{RF,C}}$ ignoring the phase shifter constraint for a nonsingular matrix $\mathbf{A}\in\mathbb{C}^{M\times M}$. Then, the combination of $\mathbf{F}_{\mathrm{RF,UC}}\mathbf{A}$ and its compensation matrix is also an optimal unconstrained analog precoder for any nonsingular matrix $\mathbf{A}$.

Algorithm 1: Find $\mathbf{F}_{\mathrm{RF,C}}$.

Input: $\mathbf{F}_{\mathrm{RF,UC}}$
Initialization: $\mathbf{F}_{(0)} = \frac{1}{\sqrt{N}} \angle\left(\mathbf{F}_{\mathrm{RF,UC}}\right)$, $n = 0$
repeat
$n \leftarrow n + 1$
$\mathbf{F}_{(n)} = \frac{1}{\sqrt{N}} \angle\left(\mathbf{F}_{\mathrm{RF,UC}} \mathbf{F}_{\mathrm{RF,UC}}^{*} \mathbf{F}_{(n-1)}\right)$
until $\left\|\mathbf{F}_{\mathrm{RF,UC}} \mathbf{F}_{\mathrm{RF,UC}}^{*} \mathbf{F}_{(n-1)} - \mathbf{F}_{(n)}\right\|_F$ converges
Output: $\mathbf{F}_{\mathrm{RF,C}} = \mathbf{F}_{(n)}$

Proof: See Appendix C. ■

By using Proposition 3, $\mathbf{F}_{\mathrm{RF,UC}}$ in (16) can be replaced by $\mathbf{F}_{\mathrm{RF,UC}}\mathbf{A}$ for any nonsingular matrix $\mathbf{A}$ without performance loss. The modified optimization problem becomes

$$\min_{\mathbf{F}_{\mathrm{RF,C}}, |[\mathbf{F}_{\mathrm{RF,C}}]_{i,j}| = \frac{1}{\sqrt{N}}, \mathbf{A}} \left\|\mathbf{F}_{\mathrm{RF,UC}}\mathbf{A} - \mathbf{F}_{\mathrm{RF,C}}\right\|_{\mathrm{F}}^{2}. \tag{19}$$

Thanks to the increased degrees of freedom in the design, the constrained analog precoder $\mathbf{F}_{\mathrm{RF,C}}$ can be made closer to the optimal constrained analog precoder. The solution to (19) can be obtained by alternating minimization [39], [40]. Given a fixed $\mathbf{F}_{\mathrm{RF,C}}$, the optimal $\mathbf{A}$ is given by

$$\begin{aligned}\mathbf{A}^{(\mathrm{opt})} &= \arg\min_{\mathbf{A}} \left\|\mathbf{F}_{\mathrm{RF,UC}}\mathbf{A} - \mathbf{F}_{\mathrm{RF,C}}\right\|_{\mathrm{F}}^{2} \\ &= \mathbf{F}_{\mathrm{RF,UC}}^{*}\mathbf{F}_{\mathrm{RF,C}}. \end{aligned} \tag{20}$$

Then, assuming that $\mathbf{A}$ is fixed, the optimal $\mathbf{F}_{\mathrm{RF,C}}$ is

$$\begin{aligned}\mathbf{F}_{\mathrm{RF,C}}^{(\mathrm{opt})} &= \arg\min_{\mathbf{F}_{\mathrm{RF,C}}, |[\mathbf{F}_{\mathrm{RF,C}}]_{i,j}| = \frac{1}{\sqrt{N}}} \left\|\mathbf{F}_{\mathrm{RF,UC}}\mathbf{A} - \mathbf{F}_{\mathrm{RF,C}}\right\|_{\mathrm{F}}^{2} \\ &= \frac{1}{\sqrt{N}} \angle\left(\mathbf{F}_{\mathrm{RF,UC}}\mathbf{A}\right). \end{aligned} \tag{21}$$

Using (20) and (21), the solution can be obtained from an iterative algorithm described in Algorithm 1. The convergence of the alternating minimization algorithm is provided in [39].

Once the optimal $\mathbf{F}_{\mathrm{RF,C}}$ is found, the compensation matrix $\mathbf{F}_{\mathrm{CM}}$ is obtained from $\mathbf{F}_{\mathrm{RF,C}}$ using (17). The overall baseband precoding in the constrained case is

$$\mathbf{F}_{\mathrm{BB,C}} = \mathbf{F}_{\mathrm{CM}}\mathbf{F}_{\mathrm{RZF}}, \tag{22}$$

where $\mathbf{F}_{\mathrm{RZF}}$ is a RZF precoder with respect to the effective channel $\mathbf{H}_{\mathrm{eff,c}} = \mathbf{H}^{*}\mathbf{F}_{\mathrm{RF,C}}\mathbf{F}_{\mathrm{CM}}$ as

$$\mathbf{F}_{\mathrm{RZF}} = \left(\mathbf{H}_{\mathrm{eff,c}}^{*}\mathbf{H}_{\mathrm{eff,c}} + \beta\mathbf{I}\right)^{-1}\mathbf{H}_{\mathrm{eff,c}}^{*}. \tag{23}$$

Algorithm 2 summarizes the overall process for the hybrid precoding design under Constraints 1-3.

V. Asymptotic Analysis for the SLNR Loss Caused by the Hybrid Structure

In this section, we analyze the SLNR loss of the proposed hybrid precoding strategy compared to the fully digital precoder. For analytical tractability, we focus on the unconstrained analog precoding case obtained in Section III. Our analysis is still useful though since the quantization due to the phase shifters becomes negligible in the proposed constrained hybrid precoding. This will be shown in Section VI. As a measure of the loss, we use the ratio of the asymptotic SLNR averaged over U users of the hybrid precoding to that of the fully digital precoding. Similarly to the hybrid precoding case in Section III, the asymptotic SLNR of user u in the fully digital precoding case can be represented as

$$\mathrm{SLNR}_{u}^{(\mathrm{FD})} \xrightarrow{a.s.} \gamma_{u}^{(\mathrm{FD})}, \tag{24}$$

where $\gamma_{1}^{(\mathrm{FD})}, \ldots, \gamma_{U}^{(\mathrm{FD})}$ are the unique nonnegative solution of

$$\gamma_{u}^{(\mathrm{FD})} = \mathrm{Tr}\left(\mathbf{R}_{u}\left(\sum_{j=1}^{U}\frac{\mathbf{R}_{j}}{1+\gamma_{j}^{(\mathrm{FD})}} + \frac{U}{\rho}\mathbf{I}\right)^{-1}\right). \tag{25}$$

Let $\gamma_{u}^{(\mathrm{HB})}$ denote the asymptotic SLNR of the hybrid precoding in (8). Then, we define the quality performance metric as

$$\gamma_{\mathrm{H/F}} = \frac{\frac{1}{U}\sum_{u=1}^{U}\gamma_{u}^{(\mathrm{HB})}}{\frac{1}{U}\sum_{u=1}^{U}\gamma_{u}^{(\mathrm{FD})}}, \tag{26}$$

and $\gamma_{\mathrm{H/F}}$ satisfies $0 \le \gamma_{\mathrm{H/F}} \le 1$. Note that $10\log_{10}\gamma_{\mathrm{H/F}}$ indicates the average SLNR loss in dB caused by the hybrid precoding compared to the fully digital precoding. Therefore, if $\mathbf{R}_{1}, \ldots, \mathbf{R}_{U}$ are given, the SLNR loss can be calculated by using (8), (24), and (26). The SLNR loss, however, does not have a closed form due to the fixed point solutions.

In the following propositions, some special cases are introduced where the SLNR loss metric has a closed form. For the general case, we derive an approximation of the SLNR loss met-

Algorithm 2: Hybrid Precoding Design for Multiuser Massive MIMO.

Step 1: Find an unconstrained analog precoding matrix $\mathbf{F}_{\mathrm{RF,UC}}$

$$\mathbf{F}_{\mathrm{RF,UC}} = \mathbf{V}\mathbf{U}^{*},$$

where $\tilde{M} = \min(M, \mathrm{rank}(\sum_{u=1}^{U}\mathbf{R}_{u}))$, $\mathbf{V} \in \mathbb{C}^{N \times \tilde{M}}$ is composed of $\tilde{M}$ dominant eigenvectors of $\sum_{u=1}^{U}\mathbf{R}_{u}$, and $\mathbf{U} \in \mathbb{C}^{M \times \tilde{M}}$ is a semi-unitary matrix such that $\mathbf{U}^{*}\mathbf{U} = \mathbf{I}$.
Step 2: Find a constrained analog precoding matrix $\mathbf{F}_{\mathrm{RF}}$ using Algorithm 1.
Step 3: Construct a compensation matrix, $\mathbf{F}_{\mathrm{CM}}$, as

$$\mathbf{F}_{\mathrm{CM}} = \mathbf{V}_{\mathrm{RF}}\mathbf{D}_{\mathrm{RF}}^{-1}\mathbf{V}_{\mathrm{RF}}^{*},$$

where $\mathbf{F}_{\mathrm{RF,C}} = \mathbf{V}_{\mathrm{RF}}\mathbf{D}_{\mathrm{RF}}\mathbf{V}_{\mathrm{RF}}^{*}$ by SVD.
Step 4: Construct an RZF precoding matrix, $\mathbf{F}_{\mathrm{RZF}}$, as

$$\mathbf{F}_{\mathrm{RZF}} = \left(\mathbf{F}_{\mathrm{CM}}^{*}\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{H}\mathbf{H}^{*}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{CM}} + \beta\mathbf{I}\right)^{-1}\mathbf{F}_{\mathrm{CM}}^{*}\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{H}.$$

Step 5: Construct an overall baseband precoding matrix $\mathbf{F}_{\mathrm{BB,C}}$ and a overall hybrid precoding matrix $\mathbf{F}_{\mathrm{HB}}$ as

$$\mathbf{F}_{\mathrm{BB}} = \mathbf{F}_{\mathrm{CM}}\mathbf{F}_{\mathrm{RZF}},$$
$$\mathbf{F}_{\mathrm{HB}} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{CM}}\mathbf{F}_{\mathrm{RZF}}.$$

ric in Proposition 6. We assume that $\frac{U}{N}$ and $\frac{M}{N}$ have constant values as N goes to infinity. Note that $0 < U \leq M \leq N$.

Proposition 4: For uncorrelated channels, i.e. $\mathbf{R}_u = \mathbf{I}, \forall u$, the SLNR loss metric $\gamma_{\mathrm{H/F}}$ becomes

$$\gamma_{\mathrm{H/F}} = \frac{((M-U)\rho - U) + \sqrt{((M-U)\rho - U)^2 + 4MU\rho}}{((N-U)\rho - U) + \sqrt{((N-U)\rho - U)^2 + 4NU\rho}}. \tag{27}$$

Proof: See Appendix D. ■

As $\rho \to \infty$, the limit of $\gamma_{\mathrm{H/F}}$ in (47) becomes

$$\lim_{\rho \to \infty} \gamma_{\mathrm{H/F}} = \frac{M-U}{N-U}. \tag{28}$$

This indicates that, in the high SNR region ($\rho \to \infty$) for the uncorrelated channels, the SLNR loss caused by the hybrid precoding becomes severe if $M \ll N$ and U approaches M, i.e., $U \approx M \ll N$.

In the next proposition, we show that the SLNR loss decreases as the channels become more spatially correlated. In this correlated case, the covariance matrix $\mathbf{R}_u$ is likely to be ill-conditioned, i.e., the eigenvalues are not evenly distributed, and a few dominant eigenvalues account for most of the sum of all the eigenvalues. The following proposition shows the extreme case where there is no SLNR loss from the hybrid precoding in the correlated channels.

Proposition 5: For correlated channels, if $\sum_{u=1}^{U} \mathbf{R}_u$ is rank-deficient and its rank is lower than or equal to M, then the SLNR loss metric $\gamma_{\mathrm{H/F}}$ is equal to one. In other words, hybrid precoding has the same asymptotic SLNR as that of the fully digital precoding.

Proof: See Appendix E. ■

Now let us consider a general correlated channel where the rank of $\mathbf{R}_{\mathrm{tot}}$ is not strictly less than M. Although the $(N-M)$ smallest eigenvalues are not exactly zero, it is possible for those eigenvalues to become much smaller than the other dominant eigenvalues in the highly correlated channels. It is intuitive that the smaller those non-dominant eigenvalues are, the smaller the loss from the hybrid precoding. A question still remains about how much the exact loss will be according to the portions of the small eigenvalues. For quantitative analysis, let $\lambda_1, \ldots, \lambda_N$ be the nonnegative eigenvalues of $\mathbf{R}_{\mathrm{tot}}$ in descending order and define a metric, κ_{ch}, as the ratio of the sum of M largest eigenvalues to the sum of all eigenvalues, i.e., $\kappa_{\mathrm{ch}} = \frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$. This metric κ_{ch} ranging from $\frac{M}{N}$ to 1 is a metric for eigenvalue concentration. The goal of the quantitative analysis is to express the SLNR loss $\gamma_{\mathrm{H/F}}$ as a function of the concentration metric κ_{ch} and other system parameters such as $\frac{U}{N}$ and $\frac{M}{N}$, which can provide a useful insight to the relation between both metrics.

The closed form expressions on the SLNR metric $\gamma_{\mathrm{H/F}}$ in Proposition 4 and 5 are the special cases when $\kappa_{\mathrm{ch}} = \frac{M}{N}$ and $\kappa_{\mathrm{ch}} = 1$, respectively. The SLNR loss metric $\gamma_{\mathrm{H/F}}$, however, does not have a closed form expression if $\frac{M}{N} < \kappa_{\mathrm{ch}} < 1$. Instead of pursuing exact expressions, we find an approximate expression of $\gamma_{\mathrm{H/F}}$ resorting to two assumptions to get insight to the impact of κ_{ch} and other parameters on $\gamma_{\mathrm{H/F}}$. First, all users' deterministic SLNR's are the same as the average value as used previously. Second, all the M largest eigenvalues have an identical value that is their average values as

$$\begin{aligned} \bar{\lambda}_{\mathrm{L}} &= \frac{1}{M} \sum_{m=1}^{M} \lambda_m \\ &= \frac{\kappa_{\mathrm{ch}}}{M} \mathrm{Tr}(\mathbf{R}_{\mathrm{tot}}), \end{aligned} \tag{29}$$

and the $N-M$ remaining eigenvalues have the same value as

$$\begin{aligned} \bar{\lambda}_{\mathrm{S}} &= \frac{1}{N-M} \sum_{n=M+1}^{N} \lambda_n \\ &= \frac{1-\kappa_{\mathrm{ch}}}{N-M} \mathrm{Tr}(\mathbf{R}_{\mathrm{tot}}). \end{aligned} \tag{30}$$

From (13) and the above assumptions, the deterministic SLNR of the hybrid precoding and the fully digital precoding become the nonnegative unique solution of

$$\begin{aligned} \gamma^{(\mathrm{HB})} &= \frac{1}{U} \sum_{m=1}^{M} \frac{1}{\frac{1}{1+\gamma^{(\mathrm{HB})}} + \frac{1}{\rho \bar{\lambda}_m}} \\ &= \frac{1}{U} \frac{M}{\frac{1}{1+\gamma^{(\mathrm{HB})}} + \frac{1}{\rho \bar{\lambda}_{\mathrm{L}}}}, \end{aligned} \tag{31}$$

and

$$\begin{aligned} \gamma^{(\mathrm{FD})} &= \frac{1}{U} \sum_{n=1}^{N} \frac{1}{\frac{1}{1+\gamma^{(\mathrm{FD})}} + \frac{1}{\rho \bar{\lambda}_n}} \\ &= \frac{1}{U} \frac{M}{\frac{1}{1+\gamma^{(\mathrm{FD})}} + \frac{1}{\rho \bar{\lambda}_{\mathrm{L}}}} + \frac{1}{U} \frac{N-M}{\frac{1}{1+\gamma^{(\mathrm{FD})}} + \frac{1}{\rho \bar{\lambda}_{\mathrm{S}}}}. \end{aligned} \tag{32}$$

In the following proposition, the approximate SLNR loss metric $\gamma_{\mathrm{H/F}}$ using the above two assumptions is derived in a closed form.

Proposition 6: For spatially correlated channels where $\mathrm{Tr}(\mathbf{R}_u) = N$ for all u, if the deterministic SLNR of the hybrid precoding and the fully digital precoding are given by the solutions to (31) and (32), then the SLNR loss metric $\gamma_{\mathrm{H/F}}$ becomes

$$\gamma_{\mathrm{H/F}} = \frac{\left(\frac{1}{A} - \frac{1}{B} - 1\right) + \sqrt{\left(\frac{1}{A} - \frac{1}{B} - 1\right)^2 + \frac{4}{A}}}{-6\left(D + \varphi H + \frac{G}{\varphi H}\right)^{-1} - 2}, \tag{33}$$

where $\varphi = -\frac{1}{2} + \frac{1}{2}\sqrt{3}i$ and

$$\begin{aligned} &A = \frac{U}{\rho \kappa_{\mathrm{ch}} N}, \quad B = \frac{M}{\rho \kappa_{\mathrm{ch}} N}, \quad C = \frac{N-M}{\rho(1-\kappa_{\mathrm{ch}})N}, \\ &D = B + C - 1 + \frac{N}{U}, \quad E = BC\left(\frac{\rho N + U}{U}\right) - B - C \\ &F = 2D^3 - 9DE - 27BC, \quad G = D^2 - 3E, \\ &H = \left(\frac{F + \sqrt{F^2 - 4G^3}}{2}\right)^{\frac{1}{3}}. \end{aligned} \tag{34}$$

Proof: See Appendix F. ■

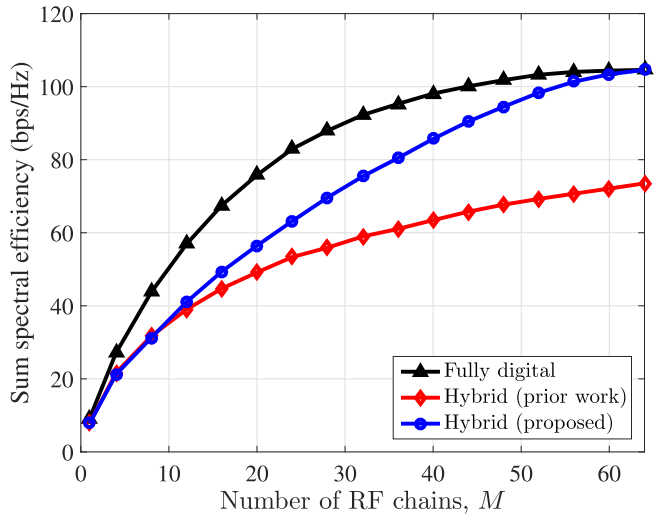


Fig. 3. Sum spectral efficiency vs. M when $U = M, N = 64, L = 5, \sigma_{\rm AS} = 10$, and SNR$= 10$ dB. Both of the hybrid precoding are performed without the phase shifter constraint. The prior work denotes the hybrid precoding technique in (2) from [16].

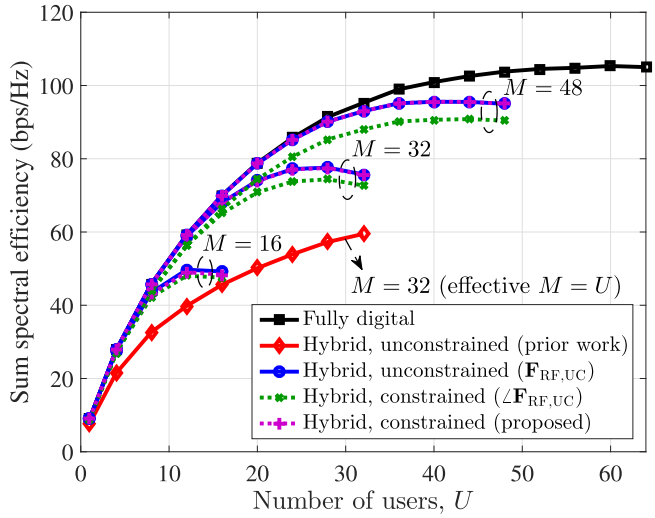


Fig. 4. Sum spectral efficiency vs. U when $M \geq U, N = 64, L = 5, \sigma_{\rm AS} = 10$, and SNR$= 10$ dB. $\mathbf{F}_{\rm RF,UC}$ is the proposed unconstrained analog precoder, and $\angle\mathbf{F}_{\rm RF,UC}$ denotes the solution of (16).

The approximate SLNR loss metric is a decreasing function with respect to $\kappa_{\rm ch}$. Since the range of $\kappa_{\rm ch}$ is $\frac{M}{N} \leq \kappa_{\rm ch} \leq 1$, the metric has a minimum value of (27) when $\kappa_{\rm ch} = \frac{M}{N}$ (uncorrelated channels), and a maximum value of one when $\kappa_{\rm ch} = 1$ (correlated channels with $\mathrm{rank}(\mathbf{R}_{\rm tot}) = M$). The approximate SLNR loss metric also depends on three other factors: $\frac{M}{N}$, $\frac{U}{N}$, and ρ $(= \frac{P_{\rm tx}}{\sigma^2})$. The dependence of the approximate SLNR loss on these factors will be discussed in Section VI as well as the validation of the approximation.

VI. Simulation Results

In this section, we use simulation results to evaluate the proposed hybrid precoding strategy. We adopt a simple geometry-based channel model for simulations. We assume that there are L channel paths between a BS and a user, and the BS has a uniform linear array with 0.5λ antenna spacing where λ is the wavelength. Let α_ℓ and ϕ_ℓ denote the ℓ-th complex path gain and angle of departure (AoD), respectively. Then, the channel vector of user u can be expressed as $\mathbf{h}_u = \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}(\phi_\ell)$ where $\mathbf{a}(\phi_\ell)$ is the array response vector given by

$$\mathbf{a}(\phi_\ell) = \begin{bmatrix} 1 & e^{j\pi \sin(\phi_\ell)} & \cdots & e^{j\pi(N-1)\sin(\phi_\ell)} \end{bmatrix}^T. \quad (35)$$

In the simulations, we assume that ϕ_ℓ's are Laplacian distributed with an angle spread $\sigma_{\rm AS}$, and $\alpha_\ell \sim \mathcal{CN}(0, \sigma_{\alpha_\ell}^2)$ where $\sigma_{\alpha_\ell}^2$'s are randomly generated from an exponential distribution and normalized such that $\sum_{\ell=1}^{L} \sigma_{\alpha_\ell}^2 = 1$.

Fig. 3 compares the proposed hybrid precoding with the prior technique in (2) when $N = 64$, $L = 5$, $\sigma_{\rm AS} = 10$, and SNR$=$ 10 dB. Since prior wok is targeted at the case of $M = U$, both are compared in that case for a fair comparison, and simulations are performed without phase shifter constraints. Fig. 3 shows that our approach outperforms prior work as U and M are not small. In addition, prior work does not approach the fully digital precoding case even when M becomes N.

In Fig. 4, the case of $U \leq M$ is evaluated, and the proposed constrained hybrid precoding under the phase shifter constraint is compared to the unconstrained case. Fig. 4 shows that the proposed constrained analog precoder combined with the compensation matrix achieves almost the same sum spectral efficiency as the unconstrained case, and outperforms using $\angle\mathbf{F}_{\rm RF,UC}$ in (16) whose elements are composed of the phase components of $\mathbf{F}_{\rm RF,UC}$. Since prior work in [16] where $\mathbf{F}_{\rm RF}$ is designed from (2) cannot be directly applied when $U \leq M$, the effective number of RF chains used is equal to U in the simulation. While M is a time-invariant system parameter, the number of users U varies over time. When less than M users are assigned, the approaches that assign a RF chain to a user such as (2) do not fully exploit all RF chains, thereby leading to inefficiency.

In Figs. 3 and 4, we assume that the phase shifters in the analog precoder can take continuous phase values. Since only quantized phases can be realized in the phase shifters in realistic implementations, it is worthwhile to investigate the phase quantization effect. Fig. 5 shows the impact of the phase quantization on sum spectral efficiency in the same environment as Fig. 4. We see that the sum spectral efficiency of the constrained case with only three or four bit quantization is close to that of the unconstrained case.

Fig. 6 shows sum spectral efficiency according to SNR. When U is not so large as in Fig. 6(a), the gain of the proposed technique is marginal compared to the prior work [16] when $M = U$. Both techniques, however, lead to a considerable rate loss compared to the fully digital precoding in the case of $M = U$ at high SNR. The figure shows that it is essential to exploit more RF chains than users. Thereby, the ability to support the case of $U < M$ is a significant advantage. In addition, even when

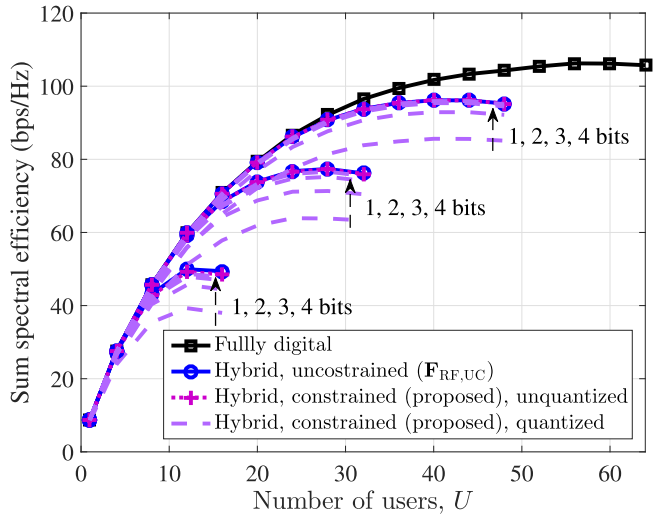
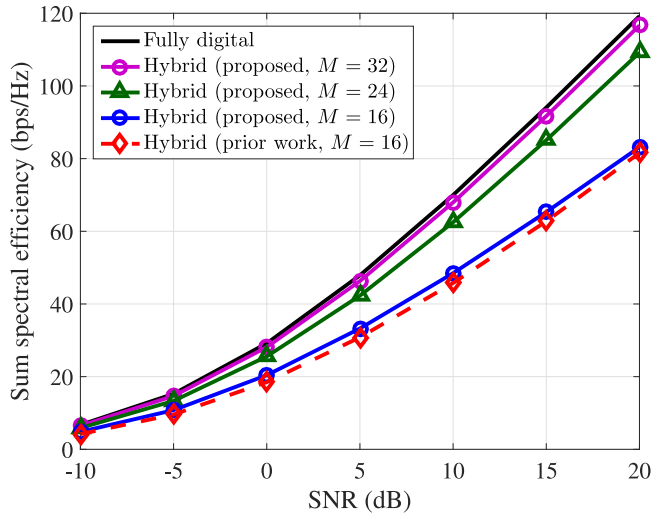
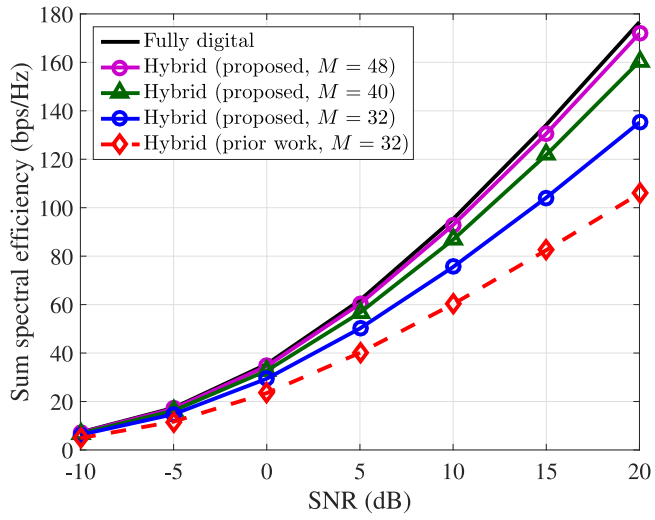


Fig. 5. Phase quantization effect on sum spectral efficiency. The simulation environment is the same as Fig. 4.



(a) $U = 16$



(b) $U = 32$

Fig. 6. Sum spectral efficiency vs. SNR when $N = 64$, $L = 5$, and $\sigma_{\mathrm{AS}} = 10$.

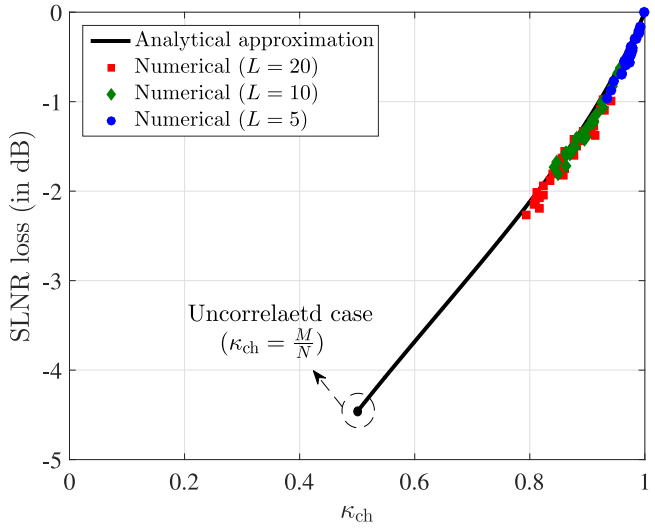


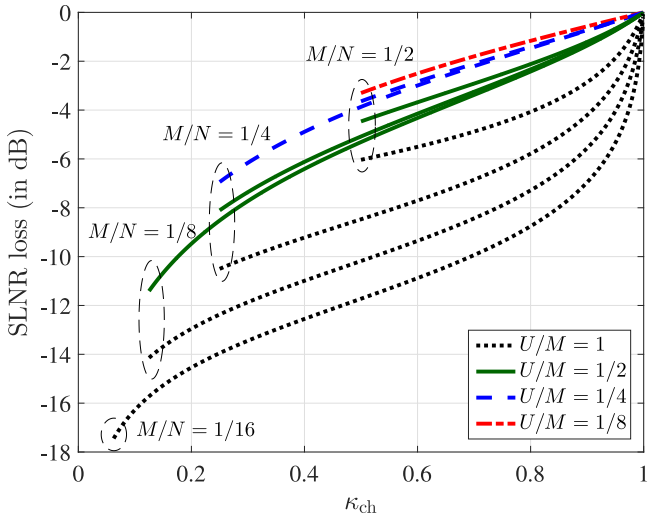
Fig. 7. SLNR gap vs. κ_{ch} when $N = 64$, $M = 32$, $U = 16$, $\sigma_{\mathrm{AS}} = 10$, and SNR$= 10$ dB.

$U = M$, the proposed technique outperforms the prior work if U is not too small as shown in Fig. 6(b).

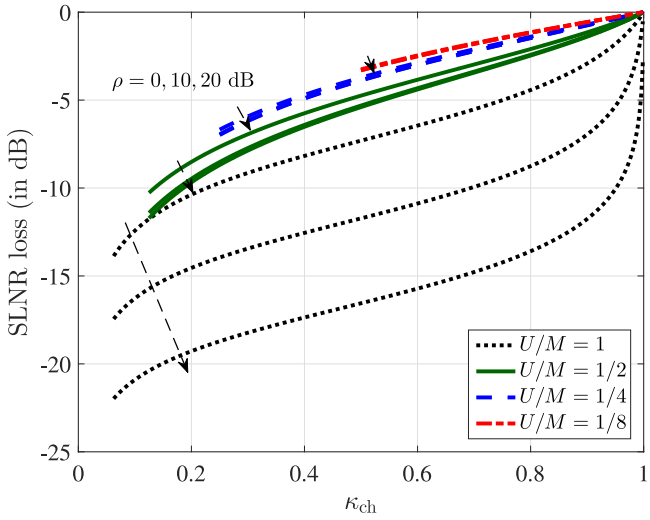
Now let us look at how much loss the proposed hybrid precoding will bring compared to the fully digital precoding. In the single-user MIMO case, the required number of RF chains to approach to the performance of the fully digital precoding depends on channel sparsity such as the number of channel paths. Since the degrees-of-freedom is limited to the number of channel paths L in the single-user MIMO case, L RF chains are enough to have the same degrees of freedom. Consequently, the loss from hybrid precoding is negligible or moderate if $M \approx L$. This is, however, not the case in multiuser MIMO systems. Another factor, U, plays an important role in the loss of hybrid precoding as shown in Section V.

Fig. 7 shows the approximate asymptotic SLNR loss derived in Section V. We can see that the analytical approximation derived in (33) are well matched to the results from numerical methods. The range of κ_{ch} is $\frac{M}{N} \leq \kappa_{\mathrm{ch}} \leq 1$ where the minimum value $\frac{M}{N}$ occurs in the uncorrelated case and the maximum value 1 occurs in the correlated case of $\mathrm{rank}(\mathbf{R}_{\mathrm{tot}}) = M$. This means that the uncorrelated channel case is the worst case in terms of hybrid architecture. While hybrid precoding results in a considerable loss in the uncorrelated channel case, the loss can be moderate in the correlated channel case. For example, the figure shows that the SLNR loss is much smaller in the spatially sparse channels than in the rich-scattering uncorrelated channels. The SLNR loss becomes smaller as the number of channel paths decreases.

Note that the approximate SLNR loss metric $\gamma_{\mathrm{H/F}}$ in (33) depends on four factors: $\frac{M}{N}$, $\frac{U}{N}$, and SNR ($\rho = \frac{P_{\mathrm{tx}}}{\sigma^2}$), and κ_{ch}. Among these factors, κ_{ch} is calculated from the users' spatial covariance matrices, which depends on the channel environment such as channel sparsity. Fig. 8 shows the relationship between the SLNR loss and κ_{ch} according to various $\frac{M}{N}$, $\frac{U}{N}$, and SNR values. In Fig. 8(a), the SNR is fixed at 10 dB, and different $\frac{M}{N}$ and $\frac{U}{M}$ values are simulated. As expected, the SLNR loss

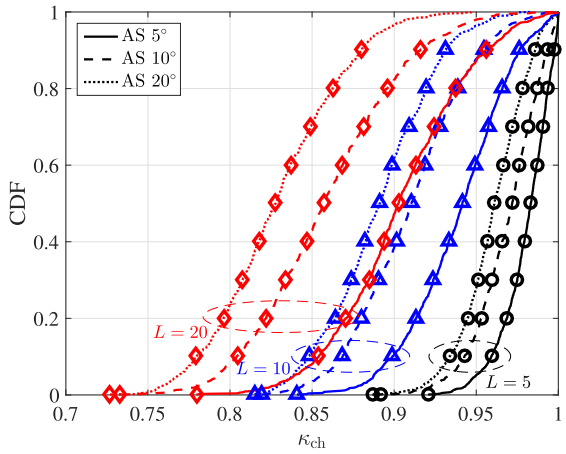


(a) ρ = 10 dB (fixed)

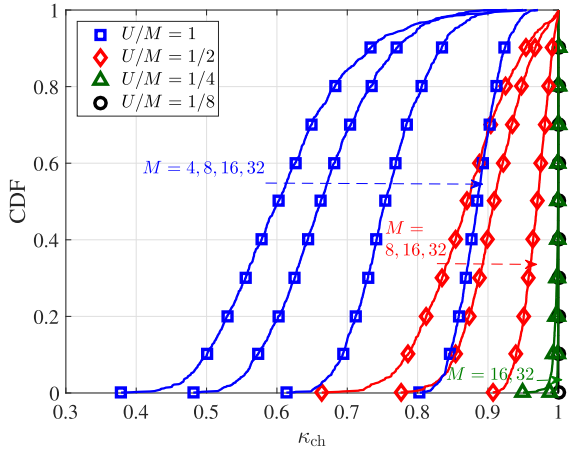


(b) $\frac{U}{N} = 1/16$ (fixed)

Fig. 8. SLNR gap (analytical) vs. κ_{ch} according to $\frac{M}{N}$, $\frac{U}{N}$, and ρ.



(a) CDF of κ_{ch} according to L and σ_{AS} when $N = 64, M = 32$, and $U = 16$.



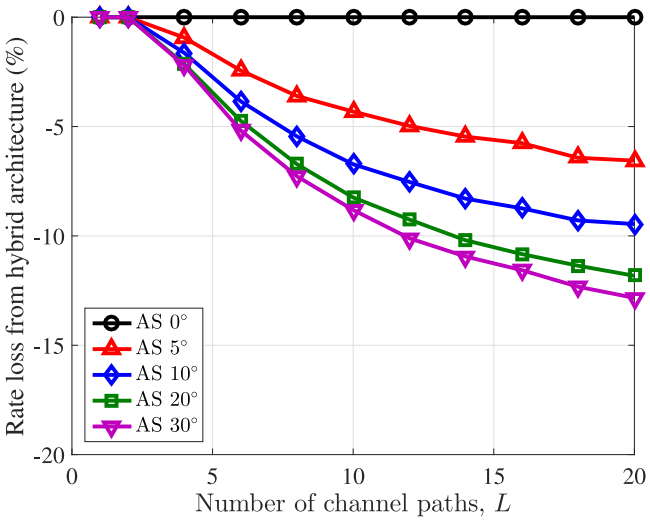
(b) CDF of κ_{ch} according to U and M when $N = 64$, $L = 5$ and $\sigma_{\mathrm{AS}} = 10°$

Fig. 9. Cumulative distribution functions of κ_{ch} according to U, M, L, and σ_{AS} when $N = 64$.

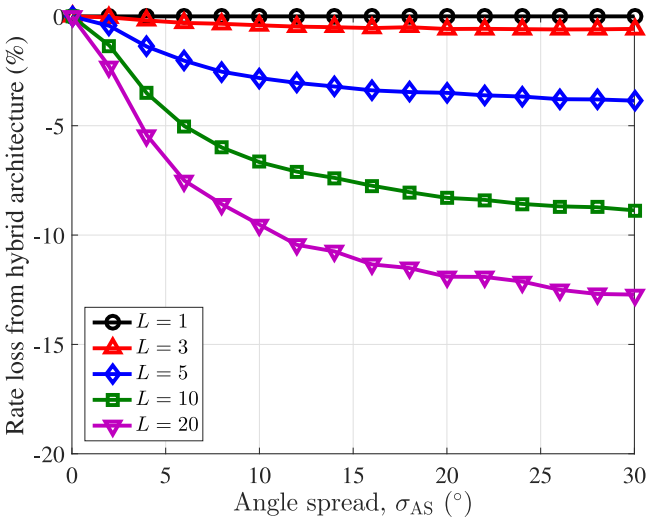
becomes small as $\frac{M}{N}$ becomes large and $\frac{U}{M}$ becomes small. One interesting point is that the SLNR loss deteriorates as $\frac{M}{N}$ decreases when $U = M$. As $\frac{U}{M}$ decreases, $\frac{M}{N}$ has less influence on the SLNR loss at the same κ_{ch} value. A similar phenomenon occurs when we look at the dependence of the SLNR loss on SNR values in Fig. 8(b). The SNR values have a significant effect on the SLNR loss only when M approaches to U. In the high SNR region, the hybrid precoding results in a disastrous SLNR loss if $M = U$. These results again emphasize the need to equip more RF chains than the expected number of users in cell deployment scenarios.

Since we have examined the relationship between the SLNR loss and κ_{ch}, let us look at how much κ_{ch} will be in various channel environments. Fig. 9 shows the cumulative distribution functions (CDFs) of κ_{ch} according to various channel and system parameters when $N = 64$. Fig. 9(a) shows that κ_{ch} has a larger value as L increases and σ_{AS} decreases, i.e., the channel becomes spatially sparser. In addition to relating to the number of channel paths, the distribution of κ_{ch} is also dependent on M and U. In Fig. 9(b), the distributions of κ_{ch} are shown according to different U and M values. We can see that κ_{ch} itself approaches to one when $\frac{U}{M}$ is small. In this small $\frac{U}{M}$ case, the dependence of κ_{ch} on M also decreases, i.e., κ_{ch} is still large even at small M values. The results in Fig. 8 and Fig. 9(b) collectively indicate that equipping more RF chains than users leads to not only a small SLNR loss at a specific κ_{ch} value but also a large κ_{ch} value itself.

Although we focus on the SLNR loss throughout this paper as the RZF is known as a precoder that maximizes the SLNR [27], it is also meaningful to look at the rate loss caused by the hybrid precoding designed based on the SLNR. Fig. 10 shows the sum rate loss according to various L and σ_{AS} values by numerical simulations. If the channel has only a single path, i.e., $L = 1$, then there is no loss from the hybrid precoding as shown in Fig. 10(b). The sum rate loss is less than 10 % as long as $L \leq 10$ when $N = 64$, $M = 32$, and $U = 16$. This rate loss



(a) Rate loss vs. L



(b) Rate loss vs. σ_{AS}

Fig. 10. Sum rate loss caused by the hybrid architecture compared with the fully digital precoding according to L and σ_{AS} when $N = 64$, $M = 32$, and $U = 16$.

can become smaller as either M becomes larger or U becomes smaller.

VII. Conclusions

In this paper, we proposed a hybrid precoding technique that uses spatial channel covariance matrices when computing analog precoders. We first obtained the unconstrained analog precoder to maximize the SLNR when the baseband precoder is RZF. Then, we used a compensation matrix that minimizes the rate loss caused by using phase shifters. Simulation results showed that the proposed hybrid precoder outperforms prior work in sum spectral efficiency, with larger gain as the number of users increases. The results also showed that the proposed constrained precoding solution combined with the compensation matrix performs close to the unconstrained case. The analysis on the SLNR loss caused by the hybrid architecture indicated that the proposed hybrid precoding is an attractive approach in highly correlated channels, e.g., when many antenna elements are packed in a small area and channels are spatially sparse.

Appendix A
Proof of Proposition 1

Let $\tilde{\mathbf{H}}^* = \mathbf{H}^*\mathbf{V}$ and $\tilde{\mathbf{h}}_u^* = \mathbf{h}_u^*\mathbf{V}$. Then, $\mathbf{h}_u^*\mathbf{W}$ in (5) becomes

$$\begin{aligned}\mathbf{h}_u^*\mathbf{W} &= \mathbf{h}_u^*\mathbf{V}\mathbf{A}\left(\mathbf{A}^*\mathbf{V}^*\mathbf{H}\mathbf{H}^*\mathbf{V}\mathbf{A} + \frac{U}{\rho}\mathbf{I}\right)^{-1}\mathbf{A}^*\mathbf{V}^* \\ &= \tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\mathbf{V}^*, \end{aligned} \tag{36}$$

where $\tilde{\mathbf{W}}_{\mathbf{A}} = (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}(\mathbf{A}\mathbf{A}^*)^{-1})^{-1}$. Using (36), the SLNR in (5) can be rewritten as

$$\begin{aligned}\mathrm{SLNR}_u &= \frac{\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\mathbf{V}^*\mathbf{h}_u\mathbf{h}_u^*\mathbf{V}\tilde{\mathbf{W}}^*\tilde{\mathbf{h}}_u^*}{\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\mathbf{V}^*\left(\mathbf{H}\mathbf{H}^* - \mathbf{h}_u\mathbf{h}_u^* + \frac{U}{\rho}\mathbf{I}\right)\mathbf{V}\tilde{\mathbf{W}}^*\tilde{\mathbf{h}}_u^*} \\ &= \frac{\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u}{\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* - \tilde{\mathbf{h}}_u\tilde{\mathbf{h}}_u^* + \frac{U}{\rho}\mathbf{I}\right)\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u} \\ &= \frac{\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u}{\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}\mathbf{I}\right)\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u - \tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u}. \end{aligned} \tag{37}$$

Let $\chi_{\mathbf{A},u}$ be defined as

$$\begin{aligned}\chi_{\mathbf{A},u} &= \frac{\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u}{\tilde{\mathbf{h}}_u^*\tilde{\mathbf{W}}_{\mathbf{A}}\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}\mathbf{I}\right)\tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u} \\ &= \frac{\tilde{\mathbf{w}}_{\mathbf{A},u}^*\tilde{\mathbf{h}}_u\tilde{\mathbf{h}}_u^*\tilde{\mathbf{w}}_{\mathbf{A},u}}{\tilde{\mathbf{w}}_{\mathbf{A},u}^*\left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}\mathbf{I}\right)\tilde{\mathbf{w}}_{\mathbf{A},u}}, \end{aligned} \tag{38}$$

where $\tilde{\mathbf{w}}_{\mathbf{A},u} = \tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u$. Then, the SLNR in (37) can be rewritten as

$$\mathrm{SLNR}_u = \frac{\chi_{\mathbf{A},u}}{1 - \chi_{\mathbf{A},u}}. \tag{39}$$

Note that $0 \le \chi_{\mathbf{A},u} < 1$ for any $\rho > 0$ and the SLNR is an increasing function of $\chi_{\mathbf{A},u}$. The optimal $\tilde{\mathbf{w}}_{\mathbf{A},u}$ that maximizes $\chi_{\mathbf{A},u}$ has the same direction as the generalized eigenvector of $(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}\mathbf{I}, \tilde{\mathbf{h}}_u\tilde{\mathbf{h}}_u^*)$. Since $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}\mathbf{I}$ is invertible, the optimal solution of $\tilde{\mathbf{w}}_{\mathbf{A},u}$ becomes

$$\begin{aligned}\tilde{\mathbf{w}}_{\mathbf{A},u}^{(\mathrm{opt})} &\propto \text{dominant eigenvector of} \left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}\mathbf{I}\right)^{-1}\tilde{\mathbf{h}}_u\tilde{\mathbf{h}}_u^* \\ &\propto \left(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}\mathbf{I}\right)^{-1}\tilde{\mathbf{h}}_u. \end{aligned} \tag{40}$$

Since $\tilde{\mathbf{w}}_{\mathbf{A},u} = \tilde{\mathbf{W}}_{\mathbf{A}}\tilde{\mathbf{h}}_u = (\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}(\mathbf{A}\mathbf{A}^*)^{-1})^{-1}\tilde{\mathbf{h}}_u$, the optimal $\tilde{\mathbf{w}}_{\mathbf{A},u}$ that maximizes the SLNR is obtained when $\mathbf{A}\mathbf{A}^* = \mathbf{I}$.

When $\mathbf{A}$ is semi-unitary, $\chi_{\mathbf{A},u}$ has a maximum value of $\chi_{\mathbf{A},u} = \tilde{\mathbf{h}}_u^*(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^* + \frac{U}{\rho}\mathbf{I})^{-1}\tilde{\mathbf{h}}_u$. Then, the SLNR in (37)

becomes

$$\mathrm{SLNR}_u = \frac{\mathbf{h}_u^* \mathbf{V} \left(\mathbf{V}^* \mathbf{H} \mathbf{H}^* \mathbf{V} + \frac{U}{\rho} \mathbf{I} \right)^{-1} \mathbf{V}^* \mathbf{h}_u}{1 - \mathbf{h}_u^* \mathbf{V} \left(\mathbf{V}^* \mathbf{H} \mathbf{H}^* \mathbf{V} + \frac{U}{\rho} \mathbf{I} \right)^{-1} \mathbf{V}^* \mathbf{h}_u}$$
$$\overset{(a)}{=} \mathbf{h}_u^* \mathbf{V} \left(\sum_{i \neq u}^{U} \mathbf{V}^* \mathbf{h}_i \mathbf{h}_i^* \mathbf{V} + \frac{U}{\rho} \mathbf{I} \right)^{-1} \mathbf{V}^* \mathbf{h}_u \quad (41)$$

where (a) comes from the matrix inversion lemma.

Appendix B
Proof of Proposition 2

Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of $\mathbf{R}_{\mathrm{tot}}$ in descending order and $\mathbf{V}_{\mathrm{A}} = [\mathbf{V}\ \mathbf{V}_0]$ be a unitary matrix such that $\mathbf{V}_0^* \mathbf{V}_0 = \mathbf{I}$ and $\mathbf{V}^* \mathbf{V}_0 = \mathbf{0}$. Since $\mathbf{V}_{\mathrm{A}}$ is a unitary matrix, $\mathbf{V}_{\mathrm{A}}^* \mathbf{R}_{\mathrm{tot}} \mathbf{V}_{\mathrm{A}}$ has the same eigenvalues as $\mathbf{R}_{\mathrm{tot}}$ and can be represented as

$$\mathbf{V}_{\mathrm{A}}^* \mathbf{R}_{\mathrm{tot}} \mathbf{V}_{\mathrm{A}} = \begin{bmatrix} \mathbf{V}^* \mathbf{R}_{\mathrm{tot}} \mathbf{V} & \mathbf{V}^* \mathbf{R}_{\mathrm{tot}} \mathbf{V}_0 \\ \mathbf{V}_0^* \mathbf{R}_{\mathrm{tot}} \mathbf{V} & \mathbf{V}_0^* \mathbf{R}_{\mathrm{tot}} \mathbf{V}_0 \end{bmatrix}. \quad (42)$$

Let the eigenvalues of $\mathbf{V}^* \mathbf{R}_{\mathrm{tot}} \mathbf{V}$ be denoted as $\nu_1 \geq \cdots \geq \nu_M$. Then, by Cauchy's interlacing theorem [41], the eigenvalues of the leading principal submatrix, $\mathbf{V}^* \mathbf{R}_{\mathrm{tot}} \mathbf{V}$, have the interlacing property

$$\lambda_{N-M+i} \leq \nu_i \leq \lambda_i, \quad \text{for } i = 1, \ldots, M. \quad (43)$$

Since $\mathbf{R}_{\mathrm{tot}}$ is Hermitian, λ_i for $i = 1, \ldots, \mathrm{rank}(\mathbf{R}_{\mathrm{tot}})$ have positive real values, and λ_i for $i > \mathrm{rank}(\mathbf{R}_{\mathrm{tot}})$ have zero values. Consequently, ν_i for $i > \mathrm{rank}(\mathbf{R}_{\mathrm{tot}})$ become zero, and

$$\lambda_i^{-1} \leq \nu_i^{-1} \leq \lambda_{N-M+i}^{-1}, \quad \text{for } i = 1, \ldots, \tilde{M}. \quad (44)$$

From (44), the constraint in (12) becomes

$$\gamma = \frac{1}{U} \sum_{m=1}^{\tilde{M}} \frac{1}{\frac{1}{1+\gamma} + \frac{1}{\rho \nu_m}} \leq \frac{1}{U} \sum_{m=1}^{\tilde{M}} \frac{1}{\frac{1}{1+\gamma} + \frac{1}{\rho \lambda_m}}, \quad (45)$$

where the equality holds if $\mathbf{V}$ is composed of $\tilde{M}$ dominant eigenvectors of $\mathbf{R}_{\mathrm{tot}}$. Since the solution of the fixed point equation with respect to γ has a maximum value if the equality holds, the proof is complete.

Appendix C
Proof of Proposition 3

Since $\mathbf{A}$ is nonsingular, $\mathbf{U}^* \mathbf{A}$ can be decomposed by SVD as $\mathbf{U}^* \mathbf{A} = \mathbf{U}_1 \mathbf{D}_1 \mathbf{V}_1^*$ where $\mathbf{U}_1 \in \mathbb{C}^{\tilde{M} \times \tilde{M}}$ is unitary and $\mathbf{D}_1 \in \mathbb{C}^{\tilde{M} \times \tilde{M}}$ is a diagonal matrix with nonzero elements. Then, $\mathbf{A}^* \mathbf{F}_{\mathrm{RF,UC}}^* \mathbf{F}_{\mathrm{RF,UC}} \mathbf{A} = \mathbf{V}_1 \mathbf{D}_1^2 \mathbf{V}_1^*$, and thus the compensation matrix with respect to $\mathbf{F}_{\mathrm{RF,UC}} \mathbf{A}$ becomes $\mathbf{V}_1 \mathbf{D}_1^{-1} \mathbf{V}_1^*$. Consequently, the combination of $\mathbf{F}_{\mathrm{RF,UC}} \mathbf{A}$ and its compensation matrix becomes

$$\mathbf{F}_{\mathrm{RF,UC}} \mathbf{A} \mathbf{F}_{\mathrm{CM}} = \mathbf{V} \mathbf{U}_1 \mathbf{D}_1 \mathbf{V}_1^* \mathbf{V}_1 \mathbf{D}_1^{-1} \mathbf{V}_1^* = \mathbf{V} \mathbf{U}_{\mathrm{alt}}^*, \quad (46)$$

where $\mathbf{U}_{\mathrm{alt}} = \mathbf{V}_1 \mathbf{U}_1^*$. Since $\mathbf{U}_{\mathrm{alt}}^* \mathbf{U}_{\mathrm{alt}} = \mathbf{U}_1 \mathbf{V}_1^* \mathbf{V}_1 \mathbf{U}_1^* = \mathbf{I}_{\tilde{M}}$, the combined matrix $\mathbf{F}_{\mathrm{RF,UC}} \mathbf{A} \mathbf{F}_{\mathrm{CM}}$ results in the same SLNR as $\mathbf{F}_{\mathrm{RF,UC}}$ by Proposition 1.

Appendix D
Proof of Proposition 4

When $\mathbf{R}_u = \mathbf{I}, \forall u$, $\gamma_u^{(\mathrm{FD})}$ in (25) is given by

$$\gamma_u^{(\mathrm{FD})} = \mathrm{Tr} \left(\left(\left(\sum_{j=1}^{U} \frac{1}{1 + \gamma_j^{(\mathrm{FD})}} + \frac{U}{\rho} \right) \mathbf{I} \right)^{-1} \right) = \frac{N}{\sum_{j=1}^{U} \frac{1}{1+\gamma_j^{(\mathrm{FD})}} + \frac{U}{\rho}}, \quad \forall u, \quad (47)$$

which implies that $\gamma_1^{(\mathrm{FD})} = \ldots = \gamma_U^{(\mathrm{FD})} = \gamma^{(\mathrm{FD})} = \frac{N}{\frac{U}{1+\gamma^{(\mathrm{FD})}} + \frac{U}{\rho}}$, and the positive solution of $\gamma^{(\mathrm{FD})}$ to this equation is equal to the denominator in (27) divided by N. In a similar way, it can be proved that $\gamma_1^{(\mathrm{HB})} = \ldots = \gamma_U^{(\mathrm{HB})} = \gamma^{(\mathrm{HB})}$ and $\gamma^{(\mathrm{HB})}$ is given by the numerator in (27) divided by N, using the fact that $\mathbf{R}_{\mathrm{tot}}$ is an identity matrix.

Appendix E
Proof of Proposition 5

Let the rank of $\mathbf{R}_{\mathrm{tot}}$ be $\tilde{M} \leq M$ and $\mathbf{V}_{\tilde{M}}$ be the eigenvector associated with its nonzero eigenvalues. Then, the rank of each user's covariance matrix $\mathbf{R}_u$ becomes at most $\tilde{M}$ and thus can be represented as $\mathbf{R}_u = \mathbf{V}_{\tilde{M}} \mathbf{Q}_u \mathbf{V}_{\tilde{M}}^*$ where $\mathbf{Q}_u \in \mathcal{C}^{\tilde{M} \times \tilde{M}}$. Note that this is not an eigenvalue decomposition, so $\mathbf{Q}_u$ is generally not a diagonal matrix. In the proposed hybrid precoding technique, the analog precoder without the phase shifter constraint is given by $\mathbf{F}_{\mathrm{RF}} = \left[\mathbf{V}_{\tilde{M}}\ \mathbf{0} \right]$, which means that only $\tilde{M}$ RF chains are used among M ones. From (9), the deterministic SLNR of user u in the hybrid precoding case is the unique nonnegative solution of

$$\gamma_u^{(\mathrm{HB})} = \mathrm{Tr} \left(\mathbf{F}_{\mathrm{RF}}^* \mathbf{R}_u \mathbf{F}_{\mathrm{RF}} \left(\sum_{j=1}^{U} \frac{\mathbf{F}_{\mathrm{RF}}^* \mathbf{R}_j \mathbf{F}_{\mathrm{RF}}}{1 + \gamma_j^{(\mathrm{HB})}} + \frac{U}{\rho} \mathbf{I}_M \right)^{-1} \right)$$
$$= \mathrm{Tr} \left(\tilde{\mathbf{Q}}_u^{(M)} \left(\sum_{j=1}^{U} \frac{\tilde{\mathbf{Q}}_j^{(M)}}{1 + \gamma_j^{(\mathrm{HB})}} + \frac{U}{\rho} \mathbf{I}_M \right)^{-1} \right)$$
$$= \mathrm{Tr} \left(\mathbf{Q}_u \left(\sum_{j=1}^{U} \frac{\mathbf{Q}_j}{1 + \gamma_j^{(\mathrm{HB})}} + \frac{U}{\rho} \mathbf{I}_{\tilde{M}} \right)^{-1} \right), \quad (48)$$

where $\tilde{\mathbf{Q}}_i^{(M)} = \left[\begin{smallmatrix} \mathbf{Q}_i & \mathbf{0}_{\tilde{M} \times (M - \tilde{M})} \\ \mathbf{0}_{(M-\tilde{M}) \times \tilde{M}} & \mathbf{0}_{(M-\tilde{M}) \times (M-\tilde{M})} \end{smallmatrix} \right]$. Let $\mathbf{V}_{\mathrm{A}} = [\mathbf{V}_{\tilde{M}}\ \mathbf{V}_{N-\tilde{M}}]$ be a unitary matrix where $\mathbf{V}_{N-\tilde{M}}$ is the null space of $\mathbf{V}_{\tilde{M}}$ such that $\mathbf{V}_{\tilde{M}}^* \mathbf{V}_{N-\tilde{M}} = \mathbf{0}_{\tilde{M} \times (N-\tilde{M})}$ and $\mathbf{V}_{N-\tilde{M}}^* \mathbf{V}_{N-\tilde{M}} = \mathbf{I}_{N-\tilde{M}}$. In the fully digital precoding case, the fixed-point

equation of the deterministic SLNR of user u in (25) can be reformulated as

$$\begin{aligned}\gamma_u^{(\mathrm{FD})} &= \mathrm{Tr}\left(\mathbf{V}_{\tilde{M}}\mathbf{Q}_u\mathbf{V}_{\tilde{M}}^*\left(\sum_{j=1}^{U}\frac{\mathbf{V}_{\tilde{M}}\mathbf{Q}_j\mathbf{V}_{\tilde{M}}^*}{1+\gamma_j^{(\mathrm{FD})}}+\frac{U}{\rho}\mathbf{I}_N\right)^{-1}\right)\\ &= \mathrm{Tr}\left(\mathbf{V}_{\mathrm{A}}\tilde{\mathbf{Q}}_u^{(N)}\mathbf{V}_{\mathrm{A}}^*\left(\sum_{j=1}^{U}\frac{\mathbf{V}_{\mathrm{A}}\tilde{\mathbf{Q}}_j^{(N)}\mathbf{V}_{\mathrm{A}}^*}{1+\gamma_j^{(\mathrm{FD})}}+\frac{U}{\rho}\mathbf{I}_N\right)^{-1}\right)\\ &= \mathrm{Tr}\left(\tilde{\mathbf{Q}}_u^{(N)}\left(\sum_{j=1}^{U}\frac{\tilde{\mathbf{Q}}_j^{(N)}}{1+\gamma_j^{(\mathrm{FD})}}+\frac{U}{\rho}\mathbf{I}_N\right)^{-1}\right)\\ &= \mathrm{Tr}\left(\mathbf{Q}_u\left(\sum_{j=1}^{U}\frac{\mathbf{Q}_j}{1+\gamma_j^{(\mathrm{FD})}}+\frac{U}{\rho}\mathbf{I}_{\tilde{M}}\right)^{-1}\right), \qquad (49)\end{aligned}$$

where $\tilde{\mathbf{Q}}_i^{(N)} = \left[\begin{smallmatrix}\mathbf{Q}_i & \mathbf{0}_{\tilde{M}\times(N-\tilde{M})}\\ \mathbf{0}_{(N-\tilde{M})\times\tilde{M}} & \mathbf{0}_{(N-\tilde{M})\times(N-\tilde{M})}\end{smallmatrix}\right]$. Since (48) is identical to (49), and the solution of these fixed point equations is unique, the proof is complete.

Appendix F
Proof of Proposition 6

The deterministic SLNR of the hybrid precoding is the nonnegative unique solution of

$$\begin{aligned}\gamma^{(\mathrm{HB})} &= \frac{1}{U}\frac{M}{\frac{1}{1+\gamma^{(\mathrm{HB})}}+\frac{1}{\rho\lambda_{\mathrm{L}}}}\\ &= \frac{M/U}{\frac{1}{1+\gamma^{(\mathrm{HB})}}+\frac{M}{\rho\kappa_{\mathrm{ch}}N}}, \qquad (50)\end{aligned}$$

and the solution is given by

$$\gamma^{(\mathrm{HB})} = \frac{\left(\frac{1}{A}-\frac{1}{B}-1\right)+\sqrt{\left(\frac{1}{A}-\frac{1}{B}-1\right)^2+\frac{4}{A}}}{2}, \qquad (51)$$

where $A=\frac{U}{\rho\kappa_{\mathrm{ch}}N}$ and $B=\frac{M}{\rho\kappa_{\mathrm{ch}}N}$. In the fully digital precoding case, the deterministic SLNR is the solution of

$$\begin{aligned}\gamma^{(\mathrm{FD})} &= \frac{1}{U}\frac{M}{\frac{1}{1+\gamma^{(\mathrm{FD})}}+\frac{1}{\rho\lambda_{\mathrm{L}}}}+\frac{1}{U}\frac{N-M}{\frac{1}{1+\gamma^{(\mathrm{FD})}}+\frac{1}{\rho\lambda_{\mathrm{S}}}}\\ &= \frac{1}{\frac{U}{M(1+\gamma^{(\mathrm{FD})})}+\frac{U}{\rho\kappa_{\mathrm{ch}}N}}+\frac{1}{\frac{U}{(N-M)(1+\gamma^{(\mathrm{FD})})}+\frac{U}{\rho(1-\kappa_{\mathrm{ch}})N}}. \qquad (52)\end{aligned}$$

Let $C=\frac{N-M}{\rho(1-\kappa_{\mathrm{ch}})N}$, $D=B+C-1+\frac{N}{U}$, $E=BC(\frac{\rho N+U}{U})-B-C$, and $x=\frac{1}{1+\gamma^{(\mathrm{FD})}}$, then (52) simplifies as

$$x^3+Dx^2+Ex-BC=0. \qquad (53)$$

Since $D\geq 0$ and $BC\geq 0$, the nonnegative solution to (53) is unique and given by

$$x=-\frac{1}{3}\left(D+\varphi H+\frac{G}{\varphi H}\right), \qquad (54)$$

where $\varphi=-\frac{1}{2}+\frac{1}{2}\sqrt{3}i$, $F=2D^3-9DE-27BC$, $G=D^2-3E$, and $H=(\frac{F+\sqrt{F^2-4G^3}}{2})^{\frac{1}{3}}$. From (51), (54), and $\gamma^{(\mathrm{FD})}=\frac{1}{x}-1$, the SLNR loss metric becomes (33).

References

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

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

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

[4] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[5] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. Inf. Theory Appl. Workshop*, Feb. 2013, pp. 1–5.

[6] X. Yu, J. Shen, J. Zhang, and K. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[7] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

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

[9] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[10] S. Han, C. I, C. Rowell, Z. Xu, S. Wang, and Z. Pan, "Large scale antenna system with hybrid digital and analog beamforming structure," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 842–847.

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

[12] R. Stirling-Gallacher and M. Rahman, "Multi-user MIMO strategies for a millimeter wave communication system using hybrid beam-forming," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2015, pp. 2437–2443.

[13] T. Bogale and L. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 4066–4071.

[14] E. Zhang and C. Huang, "On achieving optimal rate of digital precoder by RF-baseband codesign for MIMO systems," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2014, pp. 1–5.

[15] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[16] L. Liang, Y. Dai, W. Xu, and X. Dong, "How to approach zero-forcing under RF chain limitations in large mmWave multiuser systems?" in *Proc. IEEE/CIC Int. Conf. Commun. China*, Oct. 2014, pp. 518–522.

[17] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[18] J. Lee, G. T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

[19] S. Park, A. Alkhateeb, and R. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2907–2920, May 2017.

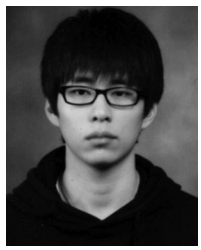
[20] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[21] E. Bjornson, D. Hammarwall, and B. Ottersten, "Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4027–4041, Oct. 2009.

[22] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4505–4515, Sep. 2014.

[23] A. Alkhateeb, R. Heath, and G. Leus, "Achievable rates of multi-user millimeter wave systems with hybrid precoding," in *Proc. IEEE Int. Conf. Commun. Workshop*, Jun. 2015, pp. 1232–1237.
[24] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
[25] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
[26] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Found. Trends Commun. Inf. Theory*, vol. 9, no. 23, pp. 113–381, 2013.
[27] M. Sadek, A. Tarighat, and A. Sayed, "Active antenna selection in multiuser MIMO communications," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1498–1510, Apr. 2007.
[28] P. Patcharamaneepakorn, S. Armour, and A. Doufexi, "On the equivalence between SLNR and MMSE precoding schemes with single-antenna receivers," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1034–1037, Jul. 2012.
[29] T. X. Tran and K. C. Teh, "Energy and spectral efficiency of leakage-based precoding for large-scale MU-MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 2041–2044, Nov. 2015.
[30] S. Park, J. Park, A. Yazdan, and R. Heath, "Optimal user loading in massive MIMO systems with regularized zero-forcing precoding," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 118–121, Feb. 2017.
[31] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007.
[32] D. Kim, G. Lee, and Y. Sung, "Two-stage beamformer design for massive MIMO downlink by trace quotient formulation," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2200–2211, Jun. 2015.
[33] A. Liu and V. Lau, "Impact of CSI knowledge on the codebook-based hybrid beamforming in massive MIMO," *IEEE Trans. Signal Process.*, vol. 64, no. 24, pp. 6545–6556, Dec. 2016.
[34] R. Méndez-Rial, N. González-Prelcic, and R. Heath, "Adaptive hybrid precoding and combining in mmWave multiuser MIMO systems based on compressed covariance estimation," in *Proc. IEEE Int. Workshop Comp. Adv. Multi-Sensor Adap. Proc.*, Dec. 2015, pp. 213–216.
[35] S. Park and R. Heath, "Spatial channel covariance estimation for mmWave hybrid MIMO architecture," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 1424–1428.
[36] R. Méndez-Rial, N. González-Prelcic, and R. Heath, "Augmented covariance estimation with a cyclic approach in DOA," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 2784–2788.
[37] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, Jul. 2012.
[38] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[39] J. A. Tropp, I. S. Dhillon, R. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005.
[40] C. Rusu, R. Méndez-Rial, N. González-Prelcic, and R. Heath, "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8380–8393, Dec. 2016.
[41] W. Haemers, "Interlacing eigenvalues and graphs," *Linear Algebra Its Appl.*, vol. 226–228, pp. 593–616, 1995.

Sungwoo Park (S'15) received the B.S. and M.S. degrees in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 2003 and 2005, respectively. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, USA. His current research interests include massive MIMO, millimeter wave communications, and compressive sensing.



Jeonghun Park (S'13) received the B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2010 and 2012, respectively. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, USA. He had worked as a graduate intern in Next Generation Standard Team, Intel, Santa Clara, CA, USA, in 2015, and as a wireless communication research intern in Connectivity Lab, Facebook, Menlo Park, CA, USA, in 2016. His research interest is developing and analyzing future wireless communication systems using tools of multi-antenna theory, information theory, and stochastic geometry.

Ali Yazdan (S'06) received the B.Sc. degree from Sharif University of Technology, Tehran, Iran, in 2004 and the M.A.Sc. degree from Simon Fraser University, Burnaby BC, Canada in 2007, both in electrical engineering, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, USA, in 2011. His research interests span topics in signal processing for mobile communications with emphasis on MIMO communications. He is currently an Engineer at Facebook's Connectivity Lab (internet.org) in Menlo Park, CA, USA.

Robert W. Heath, Jr. (S'96–M'01–SM'06–F'11) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, USA, in 1996 and 1997, respectively, and the Ph.D. degree from Stanford University, Stanford, CA, USA, in 2002, all in electrical engineering. From 1998 to 2001, he was a Senior Member of the Technical Staff then a Senior Consultant at Iospan Wireless Inc, San Jose, CA, USA, where he worked on the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication system. Since January 2002, he has been with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX, USA, where he is a Cullen Trust for Higher Education Endowed Professor, and is a member of the Wireless Networking and Communications Group. He is also the President and CEO of MIMO Wireless, Inc. He authored *Introduction to Wireless Digital Communication* (Prentice Hall, 2017), coauthored *Millimeter Wave Wireless Communications* (Prentice Hall, 2014), and authored *Digital Wireless Communication: Physical Layer Exploration Lab Using the NI USRP* (National Technology and Science Press, 2012).

He has been a coauthor of 15 award winning conference and journal papers including recently the 2010 and 2013 EURASIP *Journal on Wireless Communications and Networking* best paper awards, the 2012 Signal Processing Magazine best paper award, a 2013 Signal Processing Society best paper award, 2014 EURASIP *Journal on Advances in Signal Processing* best paper award, the 2014 Journal of Communications and Networks best paper award, the 2016 IEEE Communications Society Fred W. Ellersick Prize, and the 2016 IEEE Communications and Information Theory Societies Joint Paper Award. He was a Distinguished Lecturer in the IEEE Signal Processing Society and is an ISI Highly Cited Researcher. He is also an elected member of the Board of Governors for the IEEE Signal Processing Society, a licensed Amateur Radio Operator, a Private Pilot, and a registered Professional Engineer in Texas.