Hybrid MMSE Precoding for mmWave Multiuser MIMO Systems

Duy H. N. Nguyen†, Long Bao Le♮, and Tho Le-Ngoc♭
†Wireless Networking and Communications Group, The University of Texas at Austin, TX, USA, 78712
♭Department of Electrical and Computer Engineering, McGill University, Montréal, QC, Canada, H3A 0E9
♮INRS-EMT, Université du Québec, Montréal, QC, Canada, H3C 3P8
duy.nguyen@utexas.edu, long.le@emt.inrs.ca, tho.le-ngoc@mcgill.ca

***Abstract*—Millimeter-wave (mmWave) communication has emerged as one of the most promising technologies to deal with the increasing demand in data transmissions over wireless networks. However, due to the propagation characteristic at the mmWave band, much higher pathloss is observed compared to the commonly-used microwave band. Thus, antenna arrays become a necessary ingredient in mmWave systems because of their needed beamforming gains. Beamforming for multiple users, also known as multiuser precoding, can be utilized to further improve the spectral efficiency of mmWave systems. Unfortunately, fully digital precoding with large antenna arrays is difficult to implement due to the hardware cost and power constraint in mmWave systems. Recent works in literature have advocated the structure of hybrid analog/digital precoding for mmWave systems, in which only minor performance degradation is observed. In this work, we study hybrid precoding for multiuser mmWave systems. After reviewing recent works in literature on hybrid precoding designs, we then develop a new hybrid minimum mean-squared error (MMSE) precoder. The proposed precoder can be easily obtained by an orthogonal matching pursuit-based algorithm. Simulation results show significant performance advantages of the proposed precoder over known designs in various system settings.**

I. Introduction

Millimeter-wave (mmWave) communications have emerged as one of the most promising candidates for future cellular systems due to the significantly large and underexploited mmWave band [1]–[3]. However, antenna elements at the mmWave band usually come with much smaller aperture, which results in much lower antenna gain than that at microwave band. Thus, mmWave systems need large antenna arrays thanks to the benefit of their beamforming gains. In addition, large arrays may also allow precoding multiple data streams for multiple users, which could improve the system's spectral efficiency [4], [5]. Interestingly, packing a large number of antenna elements in a sizable space in mmWave systems is possible due to the band's short wavelength.

Multiuser precoding involves assigning the weight vectors for different mobile-stations (MS) before transmitting through the multiple antennas of the base-station (BS). Proper selection of weight vectors enables spatial separation among the users and thus supports multiplexing multiple data streams. Typically, precoding is performed at baseband by a digital signal processing (DSP) unit. However, the prohibitively high cost and power consumption of current mmWave mixed-signal hardware technologies do not allow such a transceiver architecture. Thus, mmWave systems have to rely heavily on analog or radio frequency (RF) processing [1], [5]. Analog beamforming/combining is often implemented with phase-shifters [1], which only rotate the phase of the RF signals. Recent works in precoding/combining designs for mmWave systems have advocated the use of hybrid analog/digital precoders/combiners [5]–[8]. In this hybrid structure, the analog precoder/combiner is designed to take advantage of the beamforming gains, while the digital precoder/combiner is designed to take advantage of the multiplexing gains.

Hybrid precoding/combining for single-user mmWave systems has been investigated in [5]. It was shown that hybrid precoding/combining is capable of achieving almost the same performance of the fully digital design. By taking advantage of the low-scattering property of the mmWave channel, assigning the analog precoder and combiner to the angle of departure (AoD) and angle of arrival (AoA) response vectors of most dominant channel paths is near-optimal [5]. With the obtained RF precoder/combiner, the baseband precoder/combiner then can be derived such that the resulting hybrid precoder/combiner is as close as possible to the digital one. Hybrid precoding/combining was also studied for multi-user mmWave systems [7]–[9]. In [7], [8], the authors proposed a two-stage hybrid precoding design. At the first stage, each MS and the BS jointly select a "best" combination of RF combiner and RF beamformer to maximize the channel gain to that particular MS. The baseband digital precoder is then derived as a zero-forcing (ZF) precoder by inverting the effective channel. In this work, we examine a multiuser mmWave system similar to that in [7], [8]. However, we take a different approach in deriving our proposed hybrid precoder. Specifically, while the RF combiners are decided independently at each MS, the RF precoders for all the MSs are jointly designed at the BS. The hybrid precoder is then developed with the aim of minimizing the mean-squared error (MSE) of the data streams intended for the MSs. To realize such a hybrid MMSE precoder with low computation, we then present a modified version of the orthogonal matching pursuit (OMP) algorithm [10]. Simulation results show significant performance advantage of the proposed precoder over known hybrid precoders in various system settings, including perfect AoA/AoD codebooks and quantized RF beamforming/combining codebooks.

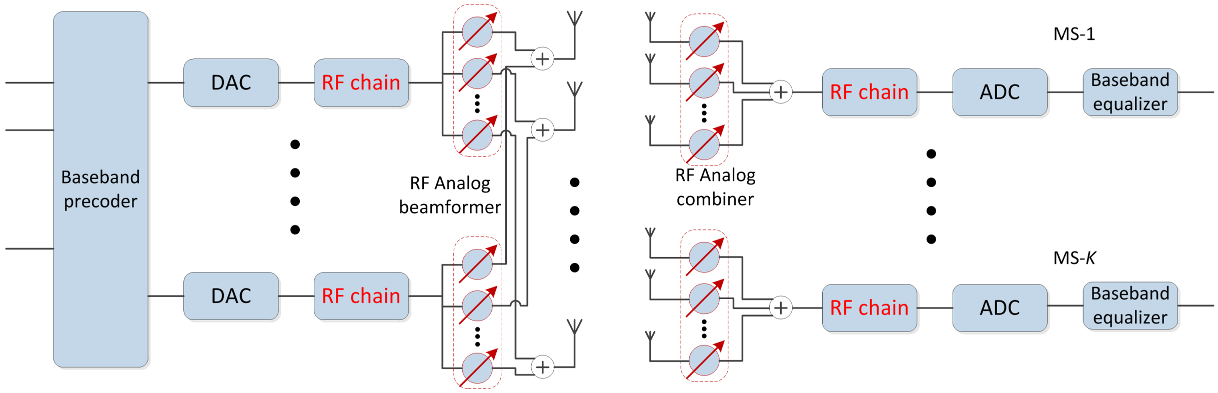

Fig. 1. Diagram of a mmWave multiuser system with hybrid analog/digital precoding and combining.

II. System Model

A. Multiuser MIMO System Model

Consider the mmWave MIMO multiuser system as illustrated in Fig. 1. A BS, equipped with M antennas and R RF chains, is communicating with K remote MSs. We assume that each MS is equipped with N receive antennas and only one RF chain. Thus, each MS can support one data stream. This assumption is justified because the implementation of mobile devices is expected to be simple, low-cost, and low-power consumption. One the other hand, the BS with much more sophisticated DSP capability, is capable of supporting multiple concurrent data streams to K MSs, if $K \leq R$.

In this paper, we focus on the downlink transmission. The BS first applies a $R \times K$ baseband precoder $\mathbf{F}_{\mathrm{B}} = [\mathbf{f}_{\mathrm{B}_1}, \ldots, \mathbf{f}_{\mathrm{B}_K}]$, where $\mathbf{f}_{\mathrm{B}_i} \in \mathbb{C}^R$ is the baseband precoding vector applied to the information symbol intended for MS-i, s_i. Following the baseband precoding and RF processing steps, the BS then applies an $M \times R$ RF precoding matrix $\mathbf{F}_{\mathrm{R}}$. Given $\mathbf{f}_i = \mathbf{F}_{\mathrm{R}}\mathbf{f}_{\mathrm{B}_i}$ as the combined BS precoding vector for MS-i, the transmitted signal is then given by

$$\mathbf{x} = \sum_{i=1}^{K} \mathbf{f}_i s_i = \mathbf{F}\mathbf{s}, \tag{1}$$

where $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_K] \in \mathbb{C}^{M \times K}$ and $\mathbf{s} = [s_1, \ldots, s_K]^T$. It is assumed that the information symbols are independent for each MS and are with unit power, i.e., $\mathbb{E}[s_i s_j] = 0$, $i \neq j$, and $\mathbb{E}[|s_i|^2] = 1$.

Denote $\mathbf{H}_i \in \mathbb{C}^{N \times M}$ as the downlink channel from the BS to MS-i, the received signal at MS-i can be modeled as

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x} + \mathbf{z}_i = \mathbf{H}_i \mathbf{f}_i s_i + \mathbf{H}_i \sum_{j \neq i}^{K} \mathbf{f}_j s_j + \mathbf{z}_i, \tag{2}$$

where the noise vector $\mathbf{z}_i$ is Gaussian distributed with zero mean and variance $\sigma^2\mathbf{I}$, i.e., $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$.

Let $\mathbf{w}_{\mathrm{R}_i}^* \in \mathbb{C}^N$ and $w_{\mathrm{B}_i}^*$ be the RF combiner and baseband equalizer, respectively, at MS-i. Denote $\mathbf{w}_i^* = \mathbf{w}_{\mathrm{R}_i}^* w_{\mathrm{B}_i}^*$ as the combined receive beamformer to process the received signal $\mathbf{y}_i$ which results in

$$\hat{s}_i = \mathbf{w}_i^H \mathbf{H}_i \mathbf{f}_i s_i + \mathbf{w}_i^H \mathbf{H}_i \sum_{j \neq i}^{K} \mathbf{f}_j s_j + \mathbf{w}_i^H \mathbf{z}_i. \tag{3}$$

It is noted that $\mathbf{F}_{\mathrm{R}}$ and $\mathbf{w}_{\mathrm{R}_i}$'s are implemented using analog phase shifter, their entries are of constant modulus. We normalize these entries to satisfy $\left|[\mathbf{F}_{\mathrm{R}}]_{m,r}\right| = \frac{1}{\sqrt{M}}$ and $\left|[\mathbf{w}_{\mathrm{R}_i}]_n\right| = \frac{1}{\sqrt{N}}, \forall i$. We denote $\mathcal{F}_{\mathrm{R}}$ and $\mathcal{W}_{\mathrm{R}}$ as the set of matrices with all constant amplitude entries, which are $\frac{1}{\sqrt{M}}$ and $\frac{1}{\sqrt{N}}$; i.e., the feasible sets of $\mathbf{F}_{\mathrm{R}}$ and $\mathbf{w}_{\mathrm{R}_i}$, respectively.

Given the received baseband signal in (3), the signal-to-interference-plus-noise ratio (SINR) at user-i is given by

$$\mathrm{SINR}_i = \frac{\left|\mathbf{w}_i^H \mathbf{H}_i \mathbf{f}_i\right|^2}{\sum_{j \neq i}^{K} \left|\mathbf{w}_i^H \mathbf{H}_i \mathbf{f}_j\right|^2 + \sigma^2 \|\mathbf{w}_i\|^2}. \tag{4}$$

Assuming Gaussian signaling is transmitted to each MS by the BS, the achievable data-rate for the transmission to MS-i is then given by $R_i = \log(1 + \mathrm{SINR}_i)$.

In this work, we are interested in jointly optimizing the baseband precoder, RF precoder, RF combiner and baseband equalizer to maximize the system sum-rate. This optimization can be stated as

$$\begin{aligned} \underset{\mathbf{F}_{\mathrm{B}}, \mathbf{F}_{\mathrm{R}}, \mathbf{w}_{\mathrm{R}_i}, w_{\mathrm{B}_i}}{\text{maximize}} \quad & \sum_{i=1}^{K} \log\left(1 + \frac{\left|\mathbf{w}_i^H \mathbf{H}_i \mathbf{f}_i\right|^2}{\sum_{j \neq i}^{K} \left|\mathbf{w}_i^H \mathbf{H}_i \mathbf{f}_j\right|^2 + \sigma^2 \|\mathbf{w}_i\|^2}\right) \\ \text{subject to} \quad & \mathbf{F}_{\mathrm{R}} \in \mathcal{F}_{\mathrm{R}} \\ & \mathbf{w}_{\mathrm{R}_i} \in \mathcal{W}_{\mathrm{R}}, \forall i \\ & \mathrm{Tr}\{\mathbf{F}_{\mathrm{R}} \mathbf{F}_{\mathrm{B}} \mathbf{F}_{\mathrm{B}}^H \mathbf{F}_{\mathrm{R}}^H\} \leq P, \end{aligned} \tag{5}$$

where P is the power constraint at the BS. In general, the optimization (5) is a nonconvex problem due to the presence of the variables $\{\mathbf{f}_i\}$ and $\{\mathbf{w}_i\}$ in the denominator of the SINR expression (4) and the multiplication of the variables. Thus, obtaining the globally optimal solution of problem (5) is not only highly complex, but also intractable for practical implementation. Instead, by taking advantage of the channel characteristics in the mmWave propagation environment as presented in the following section, we then propose low-complexity, yet efficient algorithms to compute a high-performance solution to problem (5).

B. mmWave Channel Model

One of the main characteristics of the mmWave channel is the limited number of scatters in its propagation path. This is because mmWave signaling does not reflect well to the surrounding environment. In this work, we adopt the extended Saleh-Valenzuela geometric channel model for the considered mmWave system [5]. Specifically, the channel $\mathbf{H}_i \in \mathbb{C}^{N\times M}$ from the BS to MS-i can be modeled as

$$\mathbf{H}_i = \sqrt{\frac{MN}{L_i}} \sum_{l=1}^{L_i} \alpha_{i,l} \mathbf{a}_r(\phi^r_{i,l}, \theta^r_{i,l}) \mathbf{a}_t^H(\phi^t_{i,l}, \theta^t_{i,l}), \tag{6}$$

where L_i is the number of propagation paths, $\alpha_{i,l}$ is the complex gain of the lth path, and $(\phi^r_{i,l}, \theta^r_{i,l})$ and $(\phi^t_{i,l}, \theta^t_{i,l})$ are its (azimuth, elevation) angles of arrival and departure, respectively. Then, the vectors $\mathbf{a}_r(\phi^r_{i,l}, \theta^r_{i,l})$ and $\mathbf{a}_t(\phi^t_{i,l}, \theta^t_{i,l})$ represent the normalized receive and transmit array response vectors at (azimuth, elevation) angles of $(\phi^r_{i,l}, \theta^r_{i,l})$ and $(\phi^t_{i,l}, \theta^t_{i,l})$, respectively. Finally, $\alpha_{i,l}$ is assumed to be i.i.d. Gaussian distributed and the normalization factor $\sqrt{MN/L_i}$ is added to enforce $\mathbb{E}\{\|\mathbf{H}_i\|_F^2\} = MN$.

The channel $\mathbf{H}_i$ can be restated in a more compact form as

$$\mathbf{H}_i = \mathbf{A}_{i,r} \mathbf{D}_i \mathbf{A}_{i,t}^H, \tag{7}$$

where

- $\mathbf{A}_{i,r} = \left[\mathbf{a}_r(\phi^r_{i,1}, \theta^r_{i,1}), \ldots, \mathbf{a}_r(\phi^r_{i,L_i}, \theta^r_{i,L_i})\right]$
- $\mathbf{A}_{i,t} = \left[\mathbf{a}_t(\phi^t_{i,1}, \theta^t_{i,1}), \ldots, \mathbf{a}_t(\phi^t_{i,L_i}, \theta^t_{i,L_i})\right]$
- $\mathbf{D}_i = \mathrm{diag}\Big(\alpha_{i,1}\sqrt{MN/L_i}, \ldots, \alpha_{i,L_i}\sqrt{MN/L_i}\Big)$.

It is noted that the array response vectors $\mathbf{a}_r(\phi^r_{i,l}, \theta^r_{i,l})$ and $\mathbf{a}_t(\phi^t_{i,l}, \theta^t_{i,l})$ only depend on the transmit and receive antenna array structure. Two commonly-used antenna array structures are the uniform linear array (ULA) and the uniform planar array (UPA). While the following algorithms and results presented in this work are applicable to any antenna arrays, we use UPAs in the simulations of Section V. Irrespective of the transmit or receive antenna arrays, the array response vector for a UPA in the yz-plane with W and H elements on the y and z axes is given by

$$\mathbf{a}(\phi,\theta) = \frac{1}{\sqrt{WH}} \Big[1, \ldots, \mathrm{e}^{jkd(m\sin\phi\sin\theta + n\cos\theta)}, \ldots, \mathrm{e}^{jkd[(W-1)\sin\phi\sin\theta + (H-1)\cos\theta]}\Big], \tag{8}$$

where θ and ϕ are the azimuth and elevation angles, respectively; $k = \frac{2\pi}{\lambda}$ with λ being the wavelength of the mmWave carrier frequency, and d is the inter-element spacing.

III. Review of Hybrid Precoding Designs for mmWave MIMO Systems

In this section, we briefly review two exemplary works in hybrid precoding designs: one for single-user MIMO systems [4], [5] and one for multiuser MIMO systems [7], [8]. These designs will serve as the benchmarks for comparison to the proposed hybrid MMSE precoder in this paper.

A. Single-user Spatially Sparse Precoding Design

In pioneering works [4], [5], it has been shown that hybrid precoding can obtain a near-optimal performance to the fully digital precoding for MIMO single-user mmWave systems. By exploiting the spatial structure of mmWave channels, [5] formulated the hybrid precoding design problem as a sparse reconstruction problem of the digital precoder. Specifically, given $\mathbf{F}_{\mathrm{opt}}$ as the optimal digital precoder, the RF precoder and baseband precoder are reconstructed via an approximation:

$$\begin{aligned} \underset{\mathbf{F}_{\mathrm{R}}, \mathbf{F}_{\mathrm{B}}}{\text{minimize}} \quad & \big\|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\big\|_F \\ \text{subject to} \quad & \mathbf{F}_{\mathrm{R}} \in \{\mathbf{a}_1, \ldots, \mathbf{a}_L\} \\ & \|\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\|_F^2 = P. \end{aligned} \tag{9}$$

Herein, the first constraint is to limit the search for each column of the RF precoder within a pre-determined set of L basis vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_L\}$. This set of basis vectors can be selected collectively from the transmit array response vectors at the AoD $(\phi^t_{i,l}, \theta^t_{i,l})$ of the mmWave channel for the case of perfect AoD knowledge at the transmitter, or from a codebook of quantized RF precoding vectors formed by uniform quantization of the azimuth and elevation angles [5]. Note that the constraint of $\mathbf{F}_{\mathrm{R}}$ can be embedded directly into the objective function to obtain an equivalent optimization:

$$\begin{aligned} \underset{\mathbf{F}_{\mathrm{B}}}{\text{minimize}} \quad & \big\|\mathbf{F}_{\mathrm{opt}} - \mathbf{A}\tilde{\mathbf{F}}_{\mathrm{B}}\big\|_F \\ \text{subject to} \quad & \big\|\mathrm{diag}\big(\tilde{\mathbf{F}}_{\mathrm{B}}\tilde{\mathbf{F}}_{\mathrm{B}}^H\big)\big\|_0 = R \\ & \big\|\mathbf{A}\tilde{\mathbf{F}}_{\mathrm{B}}\big\|_F^2 = P, \end{aligned} \tag{10}$$

where $\mathbf{A} = \big[\mathbf{a}_1, \ldots, \mathbf{a}_L\big]$. Due to the sparsity constraint, no more than R rows of $\tilde{\mathbf{F}}_{\mathrm{B}}$ are non-zero. As a result, these rows constitutes the baseband precoder $\mathbf{F}_{\mathrm{B}}$ and the corresponding R columns of $\mathbf{A}$ are selected to form the RF precoding $\mathbf{F}_{\mathrm{R}}$.

To obtain a sparse reconstruction of $\mathbf{F}_{\mathrm{opt}}$, an algorithmic solution based on the OMP was proposed in [5]. For ease of referencing, this algorithm is presented in the following Algorithm 1. Note that for a given RF precoder $\mathbf{F}_{\mathrm{R}}$, the baseband precoder in step **9** of Algorithm 1 is obtained as a solution to the unconstrained least-square minimization $\big\|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\big\|_F$.

Algorithm 1: Spatially Sparse Precoding Design via OMP

1 Input: $\mathbf{F}_{\mathrm{opt}}$, $\mathbf{A}$;
2 Output: $\mathbf{F}_{\mathrm{R}}$, $\mathbf{F}_{\mathrm{B}}$;
3 $\mathbf{F}_{\mathrm{res}} = \mathbf{F}_{\mathrm{opt}}$;
4 $\mathbf{F}_{\mathrm{R}} = \text{Empty}$;
5 **for** $r \le R$ **do**
6 $\quad \mathbf{\Phi} = \mathbf{A}^H \mathbf{F}_{\mathrm{res}}$;
7 $\quad k = \arg\max_l \left[\mathbf{\Phi}\mathbf{\Phi}^H\right]_{l,l}$;
8 $\quad \mathbf{F}_{\mathrm{R}} = \left[\mathbf{F}_{\mathrm{R}} \,|\, \mathbf{A}^{(k)}\right]$;
9 $\quad \mathbf{F}_{\mathrm{B}} = \left(\mathbf{F}_{\mathrm{R}}^H \mathbf{F}_{\mathrm{R}}\right)^{-1} \mathbf{F}_{\mathrm{R}}^H \mathbf{F}_{\mathrm{opt}}$;
10 $\quad \mathbf{F}_{\mathrm{res}} = \frac{\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}}{\|\mathbf{F}_{\mathrm{opt}} - \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\|_F}$;
11 Normalize $\mathbf{F}_{\mathrm{B}} = \sqrt{P}\frac{\mathbf{F}_{\mathrm{B}}}{\|\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\|_F}$.

B. Two-stage Multiuser Hybrid ZF Precoding

In more recent works [7], [8], hybrid ZF precoding has been developed for multiuser mmWave systems. Consider a similar multiuser setting as presented in Section II-A, a two-stage algorithm was proposed in [8] to obtain the hybrid precoder. In this algorithm, the first stage accounts for finding the best RF single-user RF beamforming/combining design for each MS, say MS-i, as follows:

$$(\mathbf{f}^\star_{\mathrm{R}_i}, \mathbf{w}^\star_{\mathrm{R}_i}) = \arg \max_{\mathbf{w}_{\mathrm{R}_i} \in \mathcal{W}_i, \mathbf{f}_{\mathrm{R}_i} \in \mathcal{F}_i} \big|\mathbf{w}^H_{\mathrm{R}_i} \mathbf{H}_i \mathbf{f}_{\mathrm{R}_i}\big|, \tag{11}$$

where $\mathcal{W}_i$ and $\mathcal{F}_i$ are the codebooks of RF combiners and beamformers for MS-i, respectively. MS-i then sets $\mathbf{w}_{\mathrm{R}_i} = \mathbf{w}^{\star}_{\mathrm{R}_i}$ as its RF combiner, whereas the BS forms its RF precoding matrix as $\mathbf{F}_{\mathrm{R}} = [\mathbf{f}^{\star}_{\mathrm{R}_1}, \ldots, \mathbf{f}^{\star}_{\mathrm{R}_K}]$. Effectively, $\overline{\mathbf{h}}_i^H \triangleq \mathbf{w}^H_{\mathrm{R}_i}\mathbf{H}_i\mathbf{F}_{\mathrm{R}}$ can be regarded as the downlink channel to MS-i. The second stage of the algorithm in [8] is to form the baseband precoder as the ZF precoder, i.e., $\mathbf{F}_{\mathrm{BB}} = \overline{\mathbf{H}}^H\big(\overline{\mathbf{H}}\,\overline{\mathbf{H}}^H\big)^{-1}$, where $\overline{\mathbf{H}} = [\overline{\mathbf{h}}_1, \ldots, \overline{\mathbf{h}}_K]^H$. The baseband beamforming vector for each MS is then normalized as $\mathbf{f}_{\mathrm{B}_i} = \frac{\sqrt{P/K}\mathbf{f}_{\mathrm{B}_i}}{\|\mathbf{F}_{\mathrm{R}}\mathbf{f}_{\mathrm{B}_i}\|}$ to ensure that each MS is allocated an equal portion of the total transmit power P. If $R > K$, only K RF chains are utilized in this two-stage algorithm [8].

Remark 1: While being simple to implement, the performance of ZF precoding is poor in fully loaded systems where the number of users is equal to the number of transmit antennas. In the above two-stage algorithm, the ZF baseband precoder is designed to serve K users by using only K RF chains. Thus, this ZF baseband precoder may become the limiting factor to the system sum-rate, especially with increasing K.

IV. MMSE-based Hybrid Precoding Design with Pre-determined RF Combiners

In this section, we investigate multiuser precoding designs when the RF combiner at each MS is pre-determined. Unlike the approach mentioned in Section III-B, where the RF beamformer/combiner is obtained independently for each BS-MS link [8], our proposed technique allows a joint design of RF beamforming and baseband precoder for all the MSs. In the first stage, each MS, say MS-i, independently decides its RF combiner that maximizes the its downlink channel gain:

$$\mathbf{w}^{\star}_{\mathrm{R}_i} = \arg\max_{\mathbf{w}_{\mathrm{R}_i}\in\mathcal{W}_i} \big\|\mathbf{w}^H_{\mathrm{R}_i}\mathbf{H}_i\big\|. \tag{12}$$

Denote $\mathbf{w}^H_{\mathrm{R}_i}\mathbf{H}_i = \hat{\mathbf{h}}_i^H \in \mathbb{C}^M$ as the effective MISO channel from the BS to MS-i. In the second stage, the proposed approach accounts for optimizing the precoder through the following problem

$$\begin{aligned}\underset{\mathbf{F}_{\mathrm{R}},\mathbf{f}_{\mathrm{R}_1},\ldots,\mathbf{f}_{\mathrm{R}_K}}{\text{maximize}} \quad & \sum_{i=1}^{K}\log\left(1+\frac{\big|\hat{\mathbf{h}}_i^H\mathbf{F}_{\mathrm{R}}\mathbf{f}_{\mathrm{B}_i}\big|^2}{\sum_{j\neq i}^{K}\big|\hat{\mathbf{h}}_i^H\mathbf{F}_{\mathrm{R}}\mathbf{f}_{\mathrm{B}_j}\big|^2+\sigma^2}\right) \\ \text{subject to} \quad & \mathbf{F}_{\mathrm{R}} \in \mathcal{F}_R \\ & \mathrm{Tr}\big\{\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\mathbf{F}_{\mathrm{B}}^H\mathbf{F}_{\mathrm{R}}^H\big\} \le P.\end{aligned} \tag{13}$$

Since the baseband equalizers have no effect on the achievable SINRs, they are omitted from the above optimization. Similar to the original problem (5), the above problem is also nonconvex. To this end, we examine an MMSE-based fully digital precoder design, then propose a hybrid precoder counterpart.

A. An MMSE-based Fully Digital Precoding Design

The aim of MMSE precoding is to generate the transmit precoder which results in the received signal $\hat{\mathbf{s}} = [\hat{s}_1, \ldots, \hat{s}_K]^T$ as close as possible to the original signal $\mathbf{s}$. Denote $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_K]$ as an *unnormalized* precoder at the BS and γ as a power gain factor such that $\mathbf{F} = \sqrt{1/\gamma}\mathbf{V}$ satisfies the power constraint $\mathrm{Tr}\{\mathbf{F}\mathbf{F}^H\} \le P$ at the BS.

At the receiving end, we assume that each MS applies a simple equalizer by multiplying its baseband signal with the power scaling factor $\sqrt{\gamma}$, i.e., $w_{\mathrm{B}_i} = \sqrt{\gamma}$. Substitute w_{B_i} and $\mathbf{v}_i$'s into Equation (3), $\hat{s}_i$ is given by

$$\hat{s}_i = \hat{\mathbf{h}}_i^H\mathbf{v}_i s_i + \hat{\mathbf{h}}_i^H\sum_{j\neq i}^{K}\mathbf{v}_j s_j + \sqrt{\gamma}\mathbf{w}^H_{\mathrm{R}_i}\mathbf{z}_i. \tag{14}$$

Given the sum-MSE for K data streams as $\mathbb{E}\big\{\big\|\mathbf{s}-\hat{\mathbf{s}}\big\|^2\big\}$. The MMSE precoder then can be obtained from the following optimization

$$\begin{aligned}\underset{\mathbf{V},\gamma}{\text{minimize}} \quad & \mathbb{E}\big\{\big\|\mathbf{s}-\hat{\mathbf{s}}\big\|^2\big\} \\ \text{subject to} \quad & \mathrm{Tr}\big\{\mathbf{V}\mathbf{V}^H\big\} \le \gamma P.\end{aligned} \tag{15}$$

Since the above optimization is convex [11], [12], the optimal MMSE precoder can be obtained via standard optimization techniques and given in closed-form [11]:

$$\mathbf{V}^{\star} = \left(\hat{\mathbf{H}}^H\hat{\mathbf{H}} + \frac{K\sigma^2}{P}\mathbf{I}\right)^{-1}\hat{\mathbf{H}}^H, \tag{16}$$

where $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_K]^H$; whereas the optimal scaling factor $\gamma^{\star}$ is $\|\mathbf{V}^{\star}\|_F^2/P$. The optimal fully digital MMSE precoder, denoted as $\mathbf{F}_{\mathrm{MMSE}}$, is then given by $\mathbf{F}_{\mathrm{MMSE}} = \sqrt{1/\gamma^{\star}}\mathbf{V}^{\star}$. Based on the obtained $\mathbf{F}_{\mathrm{MMSE}}$ and a pre-determined set of RF beamforming vectors, Algorithm 1 can be applied straightforwardly to approximate a hybrid precoder. Hereafter, this hybrid precoding design will be referred to as the "Two-stage Hybrid MMSE Precoding".

B. Proposed Hybrid MMSE Precoder

In this section, we propose a new hybrid MMSE precoding structure. Instead of approximating a hybrid precoder to a known fully digital precoder in Algorithm 1, the proposed hybrid precoder aims to minimize the sum-MSE of all data streams $\mathbb{E}\{\|\mathbf{s}-\hat{\mathbf{s}}\|^2\}$. Thus, the proposed hybrid precoder can bypass the step of deriving the fully digital precoder.

Denote $\mathbf{V}_{\mathrm{B}}$ as an *unnormalized* baseband precoder and γ as a power scaling factor such that $\mathbf{F}_{\mathrm{B}} = \sqrt{1/\gamma}\mathbf{V}_{\mathrm{B}}$ satisfies the power constraint $\mathrm{Tr}\big\{\mathbf{F}_{\mathrm{R}}^H\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{B}}\mathbf{F}_{\mathrm{B}}^H\big\} \le P$. Substitute $\mathbf{V} = \mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}}$ into Equation (14), we can expand the sum-MSE cost function $\mathbb{E}\{\|\mathbf{s}-\hat{\mathbf{s}}\|^2\}$ into

$$\mathbb{E}\{\|\mathbf{s}-\hat{\mathbf{s}}\|^2\} = \big\|\mathbf{I}-\hat{\mathbf{H}}\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}}\big\|_F^2 + K\gamma\sigma^2. \tag{17}$$

A hybrid precoder, which minimizes this sum-MSE, can be obtained from the following optimization

$$\begin{aligned}\underset{\mathbf{F}_{\mathrm{R}},\mathbf{V}_{\mathrm{B}},\gamma}{\text{minimize}} \quad & \mathrm{Tr}\big\{(\mathbf{I}-\hat{\mathbf{H}}\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}})(\mathbf{I}-\hat{\mathbf{H}}\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}})^H\big\} + K\gamma\sigma^2 \\ \text{subject to} \quad & \mathbf{F}_{\mathrm{R}} \in \mathcal{F}_{\mathrm{R}} \\ & \mathrm{Tr}\big\{\mathbf{F}_{\mathrm{R}}^H\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}}\mathbf{V}_{\mathrm{B}}^H\big\} \le \gamma P.\end{aligned} \tag{18}$$

We note that problem (18) is nonconvex due to the multiplication of the variables $\mathbf{F}_{\mathrm{R}}$ and $\mathbf{V}_{\mathrm{B}}$. Hence, obtaining even a locally optimal solution to problem (18) may be highly complicated. However, for a known RF precoder $\mathbf{F}_{\mathrm{R}}$, we can obtain an optimal baseband precoder $\mathbf{F}_{\mathrm{B}}$ by solving the following optimization

$$\begin{aligned} \underset{\mathbf{V}_{\mathrm{B}},\gamma}{\text{minimize}} \quad & \mathrm{Tr}\big\{(\mathbf{I}-\hat{\mathbf{H}}\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}})(\mathbf{I}-\hat{\mathbf{H}}\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}})^H\big\}+K\gamma\sigma^2 \\ \text{subject to} \quad & \mathrm{Tr}\big\{\mathbf{F}_{\mathrm{R}}^H\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}}\mathbf{V}_{\mathrm{B}}^H\big\} \le \gamma P. \end{aligned} \tag{19}$$

The optimal solution to $\mathbf{V}_{\mathrm{B}}$ can be stated in closed-form [13]

$$\mathbf{V}_{\mathrm{B}}^{\star} = \left(\mathbf{F}_{\mathrm{R}}^H\hat{\mathbf{H}}^H\hat{\mathbf{H}}\mathbf{F}_{\mathrm{R}} + \frac{K\sigma^2}{P}\mathbf{F}_{\mathrm{R}}^H\mathbf{F}_{\mathrm{R}}\right)^{-1}\mathbf{F}_{\mathrm{R}}^H\hat{\mathbf{H}}^H, \tag{20}$$

whereas the scaling factor is $\gamma^{\star} = \|\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}}\|_F^2/P$. Finally, the optimal baseband precoder $\mathbf{F}_{\mathrm{B}}^{\star}$ is given by $\sqrt{1/\gamma^{\star}}\mathbf{V}_{\mathrm{B}}^{\star}$.

In order to find the RF precoder $\mathbf{F}_{\mathrm{R}}$, we take a similar approach as in [5] to restrict its search within a set of L pre-determined basis vectors $\{\mathbf{a}_1,\ldots,\mathbf{a}_L\}$. Our proposed hybrid precoder is obtained from solving the optimization

$$\begin{aligned} \underset{\tilde{\mathbf{V}}_{\mathrm{B}},\gamma}{\text{minimize}} \quad & \mathrm{Tr}\big\{(\mathbf{I}-\hat{\mathbf{H}}\mathbf{A}\tilde{\mathbf{V}}_{\mathrm{B}})(\mathbf{I}-\hat{\mathbf{H}}\mathbf{A}\tilde{\mathbf{V}}_{\mathrm{B}})^H\big\}+K\gamma\sigma^2 \\ \text{subject to} \quad & \big\|\mathrm{diag}\{\tilde{\mathbf{V}}_{\mathrm{B}}\tilde{\mathbf{V}}_{\mathrm{B}}^H\}\big\|_0 = R \\ & \mathrm{Tr}\big\{\mathbf{A}^H\mathbf{A}\tilde{\mathbf{V}}_{\mathrm{B}}\tilde{\mathbf{V}}_{\mathrm{B}}^H\big\} \le \gamma P, \end{aligned} \tag{21}$$

where the constraint $\mathbf{F}_{\mathrm{R}} \in \{\mathbf{a}_1,\ldots,\mathbf{a}_L\}$ is embedded into the objective function with $\mathbf{A} = [\mathbf{a}_1,\ldots,\mathbf{a}_L]$. Thanks to the sparsity constraint $\big\|\mathrm{diag}\{\tilde{\mathbf{V}}_{\mathrm{B}}\tilde{\mathbf{V}}_{\mathrm{B}}^H\}\big\|_0 = R$, no more than R rows of $\tilde{\mathbf{V}}_{\mathrm{B}}$ are non-zero. These R non-zero rows are selected to form the baseband precoder $\mathbf{F}_{\mathrm{B}}$ subject to a power scaling step, whereas the corresponding R columns of $\mathbf{A}$ are selected to form the RF precoder $\mathbf{F}_{\mathrm{R}}$. Since problem (21) resembles optimization problems usually encountered in sparse signal recovery, extensive literature on this topic can be readily used to solve it. Here, we apply the OMP algorithm [10] to obtain the proposed hybrid precoder, referred to as the "Hybrid MMSE precoding". The algorithm pseudo-code is presented in Algorithm 2, in which step **9** utilizes the baseband precoder as a solution of the MSE minimization problem (19). This is the key difference to the least-square baseband solution in Algorithm 1. In terms of complexity, Algorithm 2 does not require a pre-determined digital precoder, nor introduce additional computations, compared to Algorithm 1.

Algorithm 2: Proposed Hybrid MMSE Precoding via OMP

1. Input: $\hat{\mathbf{H}}$, $\mathbf{A}$;
2. Output: $\mathbf{F}_{\mathrm{R}}$, $\mathbf{F}_{\mathrm{B}}$;
3. $\mathbf{V}_{\mathrm{res}} = \mathbf{I}$;
4. $\mathbf{F}_{\mathrm{R}} = \mathrm{Empty}$;
5. **for** $r \le R$ **do**
6. $\quad \mathbf{\Phi} = \mathbf{A}^H\hat{\mathbf{H}}^H\mathbf{V}_{\mathrm{res}}$;
7. $\quad k = \arg\max_l \left[\mathbf{\Phi}\mathbf{\Phi}^H)\right]_{l,l}$;
8. $\quad \mathbf{F}_{\mathrm{R}} = [\mathbf{F}_{\mathrm{R}}|\mathbf{A}^{(k)}]$;
9. $\quad \mathbf{V}_{\mathrm{B}} = \left(\mathbf{F}_{\mathrm{R}}^H\hat{\mathbf{H}}^H\hat{\mathbf{H}}\mathbf{F}_{\mathrm{R}} + \frac{K\sigma^2}{P}\mathbf{F}_{\mathrm{R}}^H\mathbf{F}_{\mathrm{R}}\right)^{-1}\mathbf{F}_{\mathrm{R}}^H\hat{\mathbf{H}}^H$;
10. $\quad \mathbf{V}_{\mathrm{res}} = \frac{\mathbf{I}-\hat{\mathbf{H}}\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}}}{\|\mathbf{I}-\hat{\mathbf{H}}\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}}\|_F}$;
11. $\gamma = \frac{\mathrm{Tr}\{\mathbf{F}_{\mathrm{R}}^H\mathbf{F}_{\mathrm{R}}\mathbf{V}_{\mathrm{B}}\mathbf{V}_{\mathrm{B}}^H\}}{P}$;
12. $\mathbf{F}_{\mathrm{B}} = \frac{1}{\sqrt{\gamma}}\mathbf{V}_{\mathrm{B}}$;

V. Simulation Results

In this simulation results section, we illustrate the performance advantages of the proposed hybrid MMSE precoder to other hybrid precoding designs in the literature. We compare our proposed design to three other ones: i.) fully digital MMSE precoding presented in Section IV-A, ii.) two-stage hybrid MMSE precoding by approximating the digital MMSE precoder using Algorithm 1, and iii.) two-stage hybrid ZF precoding presented in Section III-B. We consider a MIMO system where the BS is equipped with 8×8 UPA ($M = 64$) and each MS is equipped with 4×4 UPA ($N = 16$). There are $K = 8$ MSs in the system, unless stated otherwise. The number of RF chains R is set to be equal to K. The channel to each user contains of 10 paths, i.e., $L_i = 10, \forall i$. All the channel path gains $\alpha_{i,l}$'s are assumed to be i.i.d. Gaussian distribution with variance σ_α^2. The azimuths are assumed to be uniformly distributed in $[0; 2\pi]$, and the AoA/AoD elevations are uniformly distributed in $[-\frac{\pi}{2}; \frac{\pi}{2}]$. The noise variance σ^2 is set at 1. The SNR in the plot is defined as $\mathrm{SNR} = \frac{P\sigma_\alpha^2}{K}$. In all simulations presented in Figs. 2, 3, and 4, the fully digital MMSE precoder provides the highest performance, which serves as the benchmarks for hybrid precoding designs.

Fig. 2 presents the achievable system sum-rate with different digital and hybrid precoders versus the SNR. For hybrid precoding designs, perfect AoD/AoA codebooks are assumed. Specifically, the BS utilizes all the columns of $\mathbf{A}_{1,t},\ldots,\mathbf{A}_{K,t}$ to find the RF beamformer, whereas MS-i utilizes the columns of $\mathbf{A}_{i,r}$ to find the "best" RF combiner. As observed from the figure, our proposed hybrid MMSE precoder surpasses the two-stage hybrid MMSE precoder. This is because the hybrid precoder obtained from Algorithm 1, while being near-optimal in single-user systems, does not necessarily perform well in multiuser systems. The performance of the proposed hybrid MMSE precoder is also superior to that of the two-stage hybrid ZF precoder. The reason is two-fold. First, MMSE precoding usually outperforms ZF precoding [11], [14]. Second, the proposed hybrid precoder jointly designs the RF precoder, instead of independently selecting each columns of the RF precoder as in the two-stage hybrid ZF precoder.

In Fig. 3, we compare the sum-rate performances of different precoding designs versus the number of users K (and the number of RF chains R with $R = K$). The SNR is set at -10dB. As displayed in the figure, the proposed hybrid MMSE precoding significantly outperforms the two-stage hybrid MMSE precoding, especially with high K, where the latter's performance tends to saturate. Interestingly, while performing comparably to the proposed hybrid MMSE precoding with low K, the performance of the two-stage hybrid ZF precoding even decreases with high K. In contrast, the performance of proposed hybrid MMSE precoding scales almost linearly with the number of users in the system.

Finally, Fig. 4 illustrates the system sum-rate versus SNR with quantized RF beamforming/combining codebooks. Herein, we use 3-bit uniform quantization of the azimuth angle and 3-bit uniform quantization of the elevation angle at the BS and each MS. The interested readers are referred to Equation (26) in [5] for the formulation of the RF beamform-

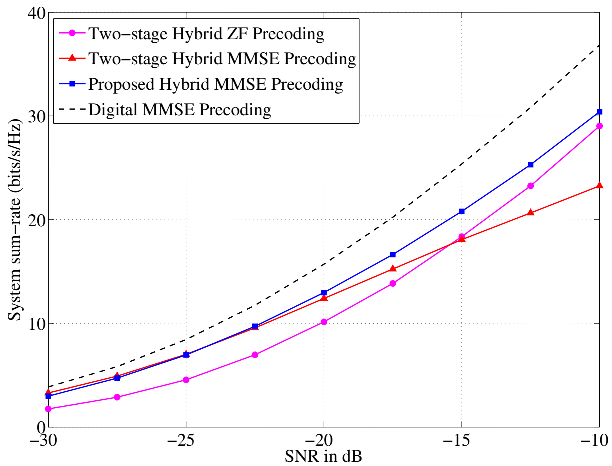

Fig. 2. System sum-rate *versus* SNR with AoD/AoA codebooks.

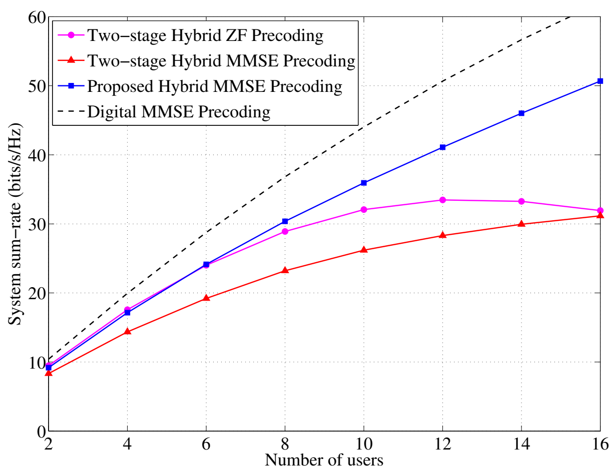

Fig. 3. System sum-rate *versus* number of users K with AoD/AoA codebooks.

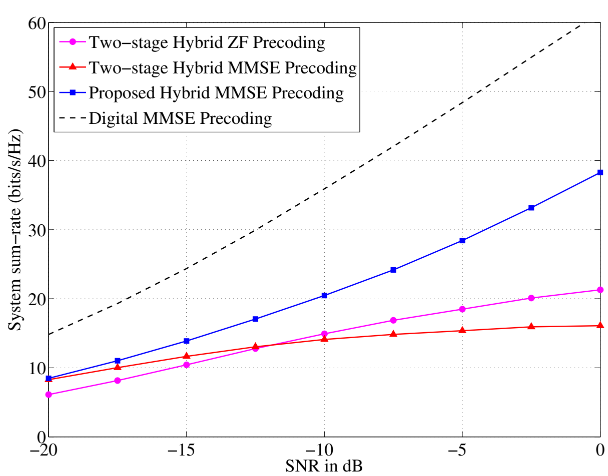

Fig. 4. System sum-rate *versus* SNR with quantized RF beamforming/combining codebooks.

ing/combining codebooks with 2^6 quantized vectors. Similar to the results presented in the previous two figures, Fig. 4 also shows a significant performance advantage of the proposed hybrid MMSE precoder. Especially at high SNR, its performance is almost double other hybrid precoding designs.

VI. Conclusion

This paper has proposed a new hybrid MMSE precoder for multiuser mmWave systems. Unlike the two-stage hybrid MMSE and ZF precoding designs, the proposed hybrid precoder aims to minimize the sum-MSE in receiving the data streams at the users. An OMP-based algorithm is then presented to obtain the proposed hybrid MMSE precoder. Simulation results show significant performance advantages of the proposed precoder over known two-stage hybrid MMSE and ZF precoders in various system settings. Our extended work in [15], involving the joint design of hybrid precoding and combining across the BS and the MSs, can further improve the system sum-rate performance over the proposed MMSE hybrid precoding design in this paper.

References

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[2] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[3] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Select. Areas in Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[4] O. El Ayach, R. W. Heath, Jr., S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. IEEE Int. Work. on Signal Process. Advances for Wireless Commun.*, Jun. 2012, pp. 100–104.

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

[6] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[7] A. Alkhateeb, R. W. Heath, Jr., and G. Leus, "Achievable rates of multiuser millimeter wave systems with hybrid precoding," in *Proc. IEEE Int. Conf. Commun.*, London, UK, Jun. 2015.

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

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

[10] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[11] M. Joham, K. Kusume, M. H. Gzara, and W. Utschick, "Transmit Wiener filter for the downlink of TDD DS-CDMA systems," in *IEEE 7th Symp. Spread-Spectrum Technol., Applicat.*, Sep. 2002, pp. 9–13.

[12] D. H. N. Nguyen and T. Le-Ngoc, "MMSE precoding for multiuser MISO downlink transmission with non-homogeneous user SNR conditions," *EURASIP J. Signal Process.*, vol. 2014:85, Jun. 2014.

[13] D. H. N. Nguyen, L. B. Le, and T. Le-Ngoc, "Multiuser MISO precoding for sum-rate maximization under multiple power constraints," in *Proc. IEEE Wireless Commun. and Networking. Conf.*, New Orleans, LA, USA, Mar. 2015, pp. 729–734.

[14] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communications - Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[15] D. H. N. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, Jr., "Hybrid MMSE precoding and combining designs for mmWave multiuser MIMO systems," *in preparation*, 2016.