

Hybrid Multi-User Precoding for mmWave Massive MIMO in Frequency-Selective Channels

Shijian Gao*, Xiang Cheng[†] and Liuqing Yang*

*Department of Electrical & Computer Engineering, Colorado State University, CO, USA.

[†]State Key Laboratory of Advanced Optical Communication Systems and Networks

Department of Electronics

School of Electronics Engineering and Computer Science, Peking University, Beijing, China.

Abstract—This paper investigates the transceiver design for downlink hybrid mmWave multi-user multi-carrier massive MIMO systems. In order to balance the processing complexity and the design flexibility, we adopt a prevalent hybrid precoding technique named hybrid block diagonalization (HBD) for downlink multi-user transmission. Aimed at maximizing the end-to-end mutual information (EEMI), a novel virtual EEMI assisted two-stage HBD scheme is judiciously devised. Apart from a low implementing complexity, the developed scheme is a more generic HBD solution, as it not only takes the frequency selectivity into account, but also removes the reliance on the high-resolution analog network. Simulations show that, even when applied with an inferior hardware configuration, the proposed HBD could still remarkably outperform its counterparts in terms of the EEMI performance at different levels of channel sparsity.

I. INTRODUCTION

To satisfy the stringent capacity requirements for the next-generation wireless systems, a well-recognized promising direction is to scale up the multiple-input multiple-output (MIMO) systems from the aspects of communication bandwidth and antenna dimension [1]–[3]. With judicious transceiver design, the resultant mmWave massive MIMO (mMIMO) system could not only enhance the link quality by large power gains, but also facilitate the spatial multiplexing via greatly refined beams [4]–[6]. By further serving multiple user equipments (UEs) simultaneously, the so-called mmWave multi-user multi-carrier massive MIMO system paves the path towards enabling massive connection in 5G cellular.

Although both academia and industry have been making remarkable efforts to design mmWave massive MIMO (mMIMO) transceivers, relevant studies in multi-user scenarios are still far from satisfaction compared to their point-to-point counterparts. Generally speaking, the transceiver design for mmWave multi-user systems encounters three prominent challenges. First, the unique hybrid structure of mmWave mMIMO imposes more complicated constraints on precoding, and thereby improving design complexity [7] [8]. Secondly, when applying orthogonal-frequency-division-multiplexing (OFDM) to mmWave mMIMO, precoding has to be jointly considered across all subcarriers for a shared analog beamformer [9] [10]. Last but not least, unlike the design in point-to-point scenarios, multi-user systems have to jointly consider user-specific signal quality variation and multi-user interferences (MUI) [11] [12].

In practice, to balance processing complexity and design flexibility, a popular option for downlink multi-user transmis-

sion is the so-termed HBD. In brief, HBD aims at eliminating MUI via hybrid precoding, such that each UE could perform independent detection within its individually MUI-free channel. To gain a broader understanding, here we make a quick review of the current representative HBD schemes. One solution proposed in [9] and [13] includes a beam-steering operation and digital pre-equalization. This leverages the channel sparsity and is thus easy-to-implement so long as a high-resolution analog network is equipped. In [12] and [14], HBD comprises of the equal-gain-transmission (EGT) analog precoding and Spencer-BD digital precoding. This solution is more generic in the sense of its independency of channel sparsity, but its applicability to frequency-selective channels has not been considered. Besides, the BS is required to work in the full-multiplexing mode. In [15] and [16], HBD is realized via preliminary beam separation and post-subspace projection. This approach works rather well if a perfect match between the multiplexing gain and the channel sparsity holds, but the performance may degrade drastically once violating this condition. Additionally, its extension to wideband channels is non-trivial as well. Although HBD devised in [17] and [18] has incorporated frequency selectivity, getting a good performance necessitates the channel sparsity and fine angular resolution. As can be seen, the effectiveness of these HBD candidates is sensitive to the hardware configuration, the operation mode, or the channel environments.

Against this background, we propose a new HBD paradigm for mmWave systems, where either side of the transceiver could employ multiple radio-frequency (RF) chains. The proposed HBD includes two stages. Specifically, at the first stage, the analog precoders will be constructed based on the criterion of maximizing “virtual” EEMI (V-EEMI). Leveraging the determined analog part, the second stage is to cancel out the residual MUIs and optimize the EEMI performance with the help of digital precoding. In addition to a perfect integration of OFDM into hybrid systems, the entire design does not impose any requirement on the angular resolution or the channel sparsity, making it a more general and appealing HBD solution to downlink multi-user transmission.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

A downlink mmWave multi-user mMIMO system is considered in this work, where N_t and N_r antennas are employed

at the BS and each of K UEs, respectively. In the studied system shown in Fig. 1, the BS communicates with each UE via N_s streams. Specifically, M_b RF chains are employed at the former satisfying $KN_s \leq M_b < N_t$, while M_u RF chains are employed at the latter satisfying $N_s \leq M_u < N_r$. Without loss of generality, we further assume $M_u = N_s$ like [15]–[17].

Similar to [19], [20], we adopt the wideband geometric channel model consisting of L dominant angular paths. Denoting D to be the total delay taps, the tap- d channel ($0 \leq d < D$) between the BS and UE- k is expressed as

$$\mathbf{H}_{k,d} = \sum_{l=1}^L \sqrt{\frac{N_t N_r}{L}} \alpha_{l,k} h(dT_s - \tau_{l,k}) \mathbf{a}_r(\theta_{l,k}) \mathbf{a}_t^*(\phi_{l,k}). \quad (1)$$

For path- l , $\alpha_{l,k} \sim \mathcal{CN}(0, 1)$ represents its amplitude; $h(\cdot)$ is the response of the pulse-shaper; $\tau_{l,k}$ is the propagation delay uniformly distributed on $[0, (N_c - 1)T_s]$; $\theta_{l,k}$ and $\phi_{l,k}$ represent the angle of arrival (AoA) and angle of departure (AoD), respectively, both uniformly distributed within $[0, 2\pi)$; $\mathbf{a}_t(\cdot)$ and $\mathbf{a}_r(\cdot)$ stand for the transmit and receive array responses, respectively. With the half-wavelength spaced uniform linear arrays (ULAs) employed, we have

$$\mathbf{a}_t(\phi) = \frac{1}{\sqrt{N_t}} [1, e^{j\pi \sin \phi}, \dots, e^{j(N_t-1)\pi \sin \phi}]^T \quad (2a)$$

$$\mathbf{a}_r(\theta) = \frac{1}{\sqrt{N_r}} [1, e^{j\pi \sin \theta}, \dots, e^{j(N_r-1)\pi \sin \theta}]^T. \quad (2b)$$

B. Input-Output Relationship

At subcarrier- n , $\mathbf{x}_n = [\mathbf{x}_{1,n}^H, \mathbf{x}_{2,n}^H, \dots, \mathbf{x}_{K,n}^H]^H$ is the input data block, with $\mathbf{x}_{k,n}$ serving for UE- k ($n \leq N, k \leq K$). All symbols in $\mathbf{x}_{k,n}$ are selected from the Gaussian constellation with $\mathbb{E}\{\mathbf{x}_{k,n} \mathbf{x}_{k,n}^H\} = \frac{1}{M_u} \mathbf{I}_{M_u}$. Before going to the antennas, $\mathbf{x}_n$ is first precoded by $\mathbf{P}_{D,n} \in \mathbb{C}^{M_b \times K M_u}$, followed by M_b N -point inverse fast Fourier transform's (IFFT's). A length- D ($D \geq N_c - 1$) cyclic-prefix (CP) will be appended to the time-domain data block before it is processed at the analog network, whose function is denoted by $\mathbf{P}_R \in \mathbb{C}^{N_t \times M_b}$. Note that, in hybrid OFDM systems, $\mathbf{P}_R$ is applied after IFFT thereby being shared by all subcarriers, so the transmitted signal at subcarrier- n can be written as

$$\mathbf{s}_n = \mathbf{P}_R \mathbf{P}_{D,n} \mathbf{x}_n. \quad (3)$$

Given the time-domain channel, the frequency-domain channel between the BS and UE- k at subcarrier- n is calculated as

$$\mathbf{H}_{k,n} = \sum_{d=0}^{D-1} \mathbf{H}_{k,d} e^{-j \frac{2\pi n}{N} d}. \quad (4)$$

Accordingly, UE- k receives the signal at subcarrier- n as

$$\mathbf{r}_{k,n} = \mathbf{H}_{k,n} \mathbf{s}_n + \boldsymbol{\eta}_{k,n} \quad (5)$$

with $\boldsymbol{\eta}_{k,n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$ representing the white Gaussian noise. Throughout this paper, the signal-to-noise ratio (SNR) is defined as $\frac{1}{\sigma^2}$.

Upon the reception of $\mathbf{r}_{k,n}$, the RF-domain signal is first combined by the analog precoder $\mathbf{W}_{R,k} \in \mathbb{C}^{N_r \times M_u}$. After removing CP and implementing M_u N -point fast Fourier transform's (FFT's), the signal goes back to frequency domain

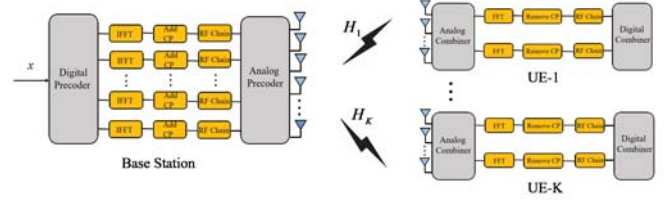


Fig. 1: The diagram of downlink mmWave multi-user system

and is combined by $\mathbf{W}_{D,k,n} \in \mathbb{C}^{M_u \times M_u}$. The I-O relationship between the BS and UE- k at subcarrier- n can be written as

$$\begin{aligned} \mathbf{y}_{k,n} &= \mathbf{W}_{D,k,n}^H \mathbf{W}_{R,k}^H \mathbf{r}_{k,n} + \mathbf{W}_{D,k,n}^H \mathbf{W}_{R,k}^H \boldsymbol{\eta}_{k,n} \\ &= \mathbf{W}_{D,k,n}^H \mathbf{W}_{R,k}^H \mathbf{H}_{k,n} \mathbf{P}_R \mathbf{P}_{D,n} \mathbf{x}_n + \boldsymbol{\xi}_{k,n} \end{aligned} \quad (6)$$

with $\boldsymbol{\xi}_{k,n} = \mathbf{W}_{D,k,n}^H \mathbf{W}_{R,k}^H \boldsymbol{\eta}_{k,n}$. By stacking all $\mathbf{y}_{k,n}$'s, we get

$$\begin{aligned} \mathbf{y}_n &= \mathbf{W}_{D,n}^H \underbrace{\begin{bmatrix} \mathbf{W}_{R,1}^H \mathbf{H}_{1,n} \mathbf{P}_R \\ \mathbf{W}_{R,2}^H \mathbf{H}_{2,n} \mathbf{P}_R \\ \vdots \\ \mathbf{W}_{R,K}^H \mathbf{H}_{K,n} \mathbf{P}_R \end{bmatrix}}_{\mathbf{H}_{e,n}} \mathbf{P}_{D,n} \mathbf{x}_n + \underbrace{\begin{bmatrix} \boldsymbol{\xi}_{1,n} \\ \boldsymbol{\xi}_{2,n} \\ \vdots \\ \boldsymbol{\xi}_{K,n} \end{bmatrix}}_{\boldsymbol{\xi}_n} \\ &= \mathbf{W}_{D,n}^H \mathbf{H}_{e,n} \sum_{k=1}^K \mathbf{P}_{D,k,n} \mathbf{x}_{k,n} + \boldsymbol{\xi}_n. \end{aligned} \quad (7)$$

where $\mathbf{P}_{D,k,n}$ is the digital precoder associated with UE- k and $\mathbf{W}_{D,n} = \text{diag}\{\mathbf{W}_{D,k,n}\}_{k=1}^K$.

C. Problem Statement

Before officially formulating the problem, a necessary condition for HBD is specifically provided below.

Definition 1: In order to achieve HBD for mmWave multi-user systems, the hybrid precoding should satisfy

$$\mathbf{H}_{k,n} \mathbf{P}_R \mathbf{P}_{D,j,n} = \mathbf{0}_{M_b \times M_u}, \quad \forall 1 \leq k \neq j \leq K. \quad (8)$$

Once Eq. (8) holds, the achievable EEMI w.r.t the channel between the BS and UE- k at subcarrier- n is calculated as [9]

$$\begin{aligned} I_{k,n} &= \det \left(\mathbf{I}_{M_u} + \frac{1}{M_u \sigma^2} (\mathbf{W}_{D,k,n}^H \mathbf{W}_{R,k}^H \mathbf{W}_{R,k} \mathbf{W}_{D,k,n})^{-1} \right. \\ &\quad \left. \mathbf{W}_{D,k,n}^H \mathbf{H}_{eff,k,n} \mathbf{H}_{eff,k,n}^H \mathbf{W}_{D,k,n} \right) \end{aligned} \quad (9)$$

with $\mathbf{H}_{eff,k,n} = \mathbf{W}_{R,k}^H \mathbf{H}_{k,n} \mathbf{P}_R \mathbf{P}_{D,k,n}$. In this paper, we assume the transceivers adopt b -bit APSs, whose adjustable angle ensemble is represented by

$$\mathcal{B} = \{0, 2\pi/2^b, \dots, 2\pi \times (2^b - 1)/2^b\}. \quad (10)$$

Targeting at maximizing EEMI, the HBD optimization problem can be formulated as

Problem Formulation 1 (PF-1) :

$$\arg \max_{\mathbf{W}_{D,n}, \mathbf{H}_{e,n}, \mathbf{P}_{D,n}} \sum_{k=1}^K \sum_{n=1}^N \log_2 I_{k,n} \quad (11a)$$

$$\text{subject to} \quad \text{Eq. (8)} \quad (11b)$$

$$\forall m, n, \mathbf{P}_R[m, n] \in \frac{1}{\sqrt{N_t}} e^{j\mathcal{B}} \quad (11c)$$

$$\forall k, p, q, \mathbf{W}_{R,k}[p, q] \in \frac{1}{\sqrt{N_r}} e^{j\mathcal{B}} \quad (11d)$$

$$\forall k, n, \|\mathbf{P}_R \mathbf{P}_{D,k,n}\| \leq M_u. \quad (11e)$$

III. HBD DESIGN

Given PF-1, we then elaborate on the proposed HBD scheme. The concept of V-EEMI is introduced initially before stepping onto the first-stage processing in RF domain and the second-stage processing in digital domain.

A. Virtual EEMI

Define the composite channel $\widetilde{\mathbf{H}}_{e,k,n}$ for UE- k as

$$\widetilde{\mathbf{H}}_{e,k,n} = [\mathbf{H}_{e,1,n}^H \cdots \mathbf{H}_{e,k-1,n}^H, \mathbf{H}_{e,k+1,n}^H \cdots \mathbf{H}_{e,K,n}^H]^H \quad (12)$$

with $\mathbf{H}_{e,k,n} = \mathbf{W}_{R,k}^H \mathbf{H}_{k,n} \mathbf{P}_R$. Definition 1 points out that $\mathbf{P}_{D,j,n}$ ($\forall j \neq k$) lies on the null-space of $\widetilde{\mathbf{H}}_{e,k,n}$, implying the dependency of $\mathbf{P}_{D,n}$ on $\mathbf{H}_{e,n}$. Denoting the concatenated transmit and receive signals to be $\mathbf{x}$ and $\mathbf{y}$, we have

$$\mathbf{y} = \underbrace{\text{diag}\{\mathbf{W}_{D,n}^H\}_{n=1}^N}_{\mathbf{W}_D^H} \underbrace{\text{diag}\{\mathbf{H}_{e,n}\}_{n=1}^N}_{\mathbf{H}_e} \underbrace{\text{diag}\{\mathbf{P}_{D,n}\}_{n=1}^N}_{\mathbf{P}_D} \mathbf{x} + \boldsymbol{\xi} \quad (13)$$

then the critical role of $\mathbf{H}_e$ for all subcarriers is apparent. We specially name $\mathbf{H}_e$ the equivalent digital channel (EDC), for it encounters the channel from the BS-end to UE-end RF chains. Unfortunately, to accurately quantify the influence of $\mathbf{H}_e$ on EEMI is still somewhat intractable. To circumvent this challenge, our idea is to derive EEMI bounds *w.r.t* $\mathbf{H}_e$, with which a proper “virtual” EEMI (V-EEMI) will be established to replace the actual yet unknown one for subsequent design.

1) *EEMI bounds*: Typically, $\mathbf{W}_{D,n}$ has to bear a block-diagonal form. But such a restriction could vanish if all UEs coordinate with each other. Follow this assumption, we perform $\text{svd}(\mathbf{H}_{e,n}) = \mathbf{U}_{e,n} \boldsymbol{\Sigma}_{e,n} \mathbf{V}_{e,n}^H$, then setting $\mathbf{P}_{D,n} = \mathbf{V}_{e,n}$ and $\mathbf{W}_{D,n} = \mathbf{U}_{e,n}$ yields

$$\mathbf{y} = \text{diag}[\boldsymbol{\Sigma}_{e,1}, \boldsymbol{\Sigma}_{e,2}, \cdots, \boldsymbol{\Sigma}_{e,N}] \mathbf{x} + \boldsymbol{\xi}. \quad (14)$$

It can be verified that the resultant $\mathbf{P}_D$ and $\mathbf{W}_D$ correspond to be the right and left singular sub-matrices of $\mathbf{H}_e$, hence the EEMI can be computed as

$$\begin{aligned} I_U &= \sum_{n=1}^N \sum_{i=1}^{KM_u} \log_2 \left(1 + \frac{\Sigma_{e,n}^2[i, i]}{M_u \sigma^2} \right) \\ &= \sum_{i=1}^{NKM_u} \log_2 \left(1 + \frac{\Sigma_e^2[i, i]}{M_u \sigma^2} \right) \end{aligned} \quad (15)$$

with $\Sigma_e^2[i, i]$ representing the i -th eigenvalue of $\mathbf{H}_e$. Apparently, Eq. (15) is the upper-bound EEMI via HBD.

Actually, no coordination exists in practice. In order for a valid HBD, a feasible choice could be setting $\mathbf{W}_{D,n} = \mathbf{I}_{KM_u}$ and $\mathbf{P}_{D,n} = \mathbf{H}_{e,n}^\dagger \boldsymbol{\Lambda}_{e,n}$ with $\boldsymbol{\Lambda}_{e,n}[i, i] = \|\mathbf{H}_{e,n}^\dagger[:, i]\|_F^{-1}$. The received signal accordingly becomes

$$\mathbf{y} = \text{diag}[\boldsymbol{\Lambda}_{e,1}, \boldsymbol{\Lambda}_{e,2}, \cdots, \boldsymbol{\Lambda}_{e,N}] \mathbf{x} + \boldsymbol{\xi}. \quad (16)$$

One can validate that $\mathbf{P}_D$ here just boils down to the zero-forcing (ZF) pre-equalizer, i.e., a special form of HBD, so the achievable EEMI is calculated as

$$\begin{aligned} I_L &= \sum_{n=1}^N \sum_{i=1}^{KM_u} \log_2 \left(1 + \frac{\|\mathbf{H}_{e,n}^\dagger[:, i]\|_F^{-2}}{M_u \sigma^2} \right) \\ &= \sum_{i=1}^{NKM_u} \log_2 \left(1 + \frac{\|\mathbf{H}_e^\dagger[:, i]\|_F^{-2}}{M_u \sigma^2} \right). \end{aligned} \quad (17)$$

Since I_L corresponds to the case where both the MUIs and inter-stream interferences are completely cancelled out, it is reasonable to regard I_L as the lower-bound EEMI via HBD.

2) *EDC design criterion*: Utilizing the derived EEMI bounds, we expect such a “virtual” EEMI (V-EEMI) falling into $[I_L, I_U]$ and being depicted by $\mathbf{H}_e$ only. If so, the RF-domain and digital-domain processing could be decoupled. Moreover, we also hope that via the guidance of V-EEMI, the RF-domain processing could be simplified.

Proposition 1: Denote $\hat{I}$ to be V-EEMI. To satisfy $I_L \leq \hat{I} \leq I_U$, it is proper to set

$$\hat{I}(\text{SNR}) = NKM_u \log_2 \left(1 + \frac{NK}{\sigma^2 \text{tr}((\mathbf{H}_e \mathbf{H}_e^H)^{-1})} \right) + \mathcal{C}(\text{SNR})$$

where $\mathcal{C}(\text{SNR})$ is an SNR-related constant.

Proof: See Appendix A. ■

Since the obtained V-EEMI is monotonically decreasing with $\sum_{n=1}^N \text{tr}((\mathbf{H}_{e,n} \mathbf{H}_{e,n}^H)^{-1})$, the V-EEMI maximization problem is equivalent to PF-2 described below.

Problem Formulation 2 (PF-2) :

$$\arg \min_{\mathbf{W}_{R,k}, \mathbf{P}_R} \sum_{n=1}^N \text{tr}((\mathbf{H}_{e,n} \mathbf{H}_{e,n}^H)^{-1}) \quad (18a)$$

$$\text{s.t.} \quad \forall m, n, \mathbf{P}_R[m, n] \in \frac{1}{\sqrt{N_r}} e^{j\mathcal{B}} \quad (18b)$$

$$\forall k, p, q, \mathbf{W}_{R,k}[p, q] \in \frac{1}{\sqrt{N_t}} e^{j\mathcal{B}} \quad (18c)$$

As expected, PF-2 only relates with $\mathbf{H}_e$, i.e., analog part. Therefore, the formidable challenge of jointly optimizing the analog and digital parts has been removed. In the next, we will detail the RF-domain processing, i.e., the first stage of HBD.

B. RF-domain processing

Note that, although PF-2 has a finite domain of definition, it is still impractical to get the minimizer of this large-scale NP-hard problem exhaustively. For tractability, we resort to entry-wise iteration for securing a local minimizer. Assuming $\mathbf{P}_R$ and $\mathbf{W}_{R,k}$ have been initialized via a certain manner¹, we

¹In practice, the initial precoders can be the random generations or the quantized version of the result in multi-beam structure. In the following simulations, this paper chooses the latter.

Algorithm 1 EWU RF-domain processing algorithm

Input: *Times*: maximum iterations; ε : terminating indicator;

Output: $\forall k, \mathbf{W}_{R,k}, \mathbf{P}_R$;

1: **Initialization:** $t = 1, e = +\infty, \mathbf{P}_R$ and $\mathbf{W}_{R,k}$.

2: Compute Eq. (19a) to get as η

3: **while** $t < \text{Times}$ and $e > \varepsilon$ **do**

4: Perform entry-wise update for $\mathbf{P}_R$ and $\mathbf{W}_R$ in sequence;

5: Compute Eq. (19a) to get $\hat{\eta}$;

6: $e = |\eta - \hat{\eta}|/|\eta|$; $\eta = \hat{\eta}, t = t + 1$;

7: **end while**

take $\mathbf{P}_R[a, b]$ as an example to clarify the subsequent update. The entry-wise updating problem is stated in PF-3.

Problem Formulation 3 (PF-3) :

$$\arg \min_{\mathbf{P}_R[a, b]} \sum_{n=1}^N \text{tr}((\mathbf{H}_{e,n} \mathbf{H}_{e,n}^H)^{-1}) \quad (19a)$$

$$\text{subject to } \mathbf{P}_R[a, b] \in \frac{1}{\sqrt{N_t}} e^{j\mathcal{B}}. \quad (19b)$$

Thanks to the simplification, the minimizer of PF-3 can be easily obtained via 2^b trials. Nevertheless, without a special treatment, even a modest-resolution APS will result in huge complexity, because the required matrix inversions are proportional to $2^b N$. Inspired by [21], [22], a simple update is going to be developed to alleviate the computational burden.

For notational simplicity, we specially define $\mathbf{P}_{R/ab} = \mathbf{P}_R$ and $\tilde{\mathbf{P}}_{R,ab} = \mathbf{0}_{N_t \times M_b}$, then replace $\mathbf{P}_{R/ab}[a, b] = 0$ and $\tilde{\mathbf{P}}_{R,ab}[a, b] = \mathbf{P}_R[a, b]$. Leveraging the Sherman-Morrison formula, the term $\text{tr}((\mathbf{H}_{e,n} \mathbf{H}_{e,n}^H)^{-1})$ can be simplified into Eq. (20) after some manipulations, with

$$\hat{\mathbf{H}}_{e,n} = [\mathbf{W}_{R,1} \mathbf{H}_{1,n}^H, \dots, \mathbf{W}_{R,K} \mathbf{H}_{K,n}^H]^H \quad (21a)$$

$$\mathbf{P}_{n/ab} = \hat{\mathbf{H}}_{e,n} (\tilde{\mathbf{P}}_{R/ab} \tilde{\mathbf{P}}_{R/ab}^H + \tilde{\mathbf{P}}_{R,ab} \tilde{\mathbf{P}}_{R,ab}^H) \hat{\mathbf{H}}_{e,n}^H \quad (21b)$$

$$\mathbf{P}_{n,ab} = \hat{\mathbf{H}}_{e,n} (\tilde{\mathbf{P}}_{R/ab} \tilde{\mathbf{P}}_{R,ab}^H + \tilde{\mathbf{P}}_{R,ab} \tilde{\mathbf{P}}_{R/ab}^H) \hat{\mathbf{H}}_{e,n}^H. \quad (21c)$$

We can verify the irrelevance between $\mathbf{P}_{n/ab}$ and $\mathbf{P}_R[a, b]$, indicating that the current update is decided by the second term of the right-hand formula in Eq. (20). Denoting this term as $\eta_{n,a,b}$, PF-3 is equivalently transformed into

Problem Formulation 4 (PF-4) :

$$\arg \min_{\mathbf{P}_R[a, b]} \sum_{n=1}^N \eta_{n,a,b} \quad (22a)$$

$$\text{subject to } \mathbf{P}_R[a, b] \in \frac{1}{\sqrt{N_t}} e^{j\mathcal{B}}. \quad (22b)$$

During each update, matrix inversions in PF-4 need to be computed once only rather than 2^b times as in PF-3. Hence the computational complexity is greatly reduced even if b is not that large. Once updating the entire RF precoder, the RF combiner can be updated similarly, so the details are omitted here. For reference, the procedures of RF-domain processing are provided in Algorithm 1.

C. Digital-domain Processing

Based on the constructed analog part, we finally come to fulfill the digital-domain processing to remove the residual MUIs and optimize the EEMI.

1) *1st-step digital-domain processing*: Due to the similarity, we elaborate on the explicit design based on subcarrier- n . For sake of presentation, the digital precoder associated with UE- k is specially decomposed as

$$\mathbf{P}_{D,k,n} = \mathbf{P}_{D,k,1,n} \mathbf{P}_{D,k,2,n} \quad (23)$$

with $\mathbf{P}_{D,k,1,n} \in \mathbb{C}^{M_b \times M_u}$ used for removing MUIs and $\mathbf{P}_{D,k,2,n} \in \mathbb{C}^{M_u \times M_u}$ used for optimizing EEMI. Performing $\text{svd}(\tilde{\mathbf{H}}_{e,k,n}) = \mathbf{U}_{e,k,n} \mathbf{\Sigma}_{e,k,n} \mathbf{V}_{e,k,n}^H$, the null-space (referring to the column space) of $\tilde{\mathbf{H}}_{e,k,n}$ is extracted as

$$\tilde{\mathbf{V}}_{e,k,n} = \mathbf{V}_{e,k,n}[:, (K-1)M_u + 1 : M_b]. \quad (24)$$

Assume $\text{rank}(\tilde{\mathbf{V}}_{e,k,n}) \geq M_u$ always holds, otherwise HBD cannot be realized anyhow. When $M_b - PM_u = 0$, the only option for $\mathbf{P}_{D,k,1,n}$ is

$$\mathbf{P}_{D,k,1,n} = \tilde{\mathbf{V}}_{e,k,n}[:, 1 : M_u], \text{rank}(\tilde{\mathbf{V}}_{e,k,n}) = M_u. \quad (25)$$

While for $M_b - PM_u > 0$, $\mathbf{P}_{D,k,1,n}$ is no longer fixed, because choosing arbitrarily M_u columns from $\tilde{\mathbf{V}}_{e,k,n}$ guarantees the MUI-free target [11]. However, a random selection may lose the power gain if the signal space has a strong correlation with the MUI-free space. To best harness the power gain while ensuring MUI-free for an intended UE, we adopt the subspace projection method proposed in [15]. Specifically, towards the channel between the UE- k and BS at subcarrier- n , denote its MUI-free space $\mathbf{P}_{k,n}^n$ and signal space $\mathbf{P}_{k,n}^s$ to be

$$\mathbf{P}_{k,n}^n = \mathbf{I} - \tilde{\mathbf{H}}_{e,k,n}^H (\tilde{\mathbf{H}}_{e,k,n} \tilde{\mathbf{H}}_{e,k,n}^H)^{-1} \tilde{\mathbf{H}}_{e,k,n} \quad (26a)$$

$$\mathbf{P}_{k,n}^s = \mathbf{H}_{e,k,n}^H (\mathbf{H}_{e,k,n} \mathbf{H}_{e,k,n}^H)^{-1} \mathbf{H}_{e,k,n}. \quad (26b)$$

Lemma 1: Let $\text{svd}(\mathbf{P}_{k,n}^s \mathbf{P}_{k,n}^n) = \bar{\mathbf{U}}_{k,n} \bar{\mathbf{\Sigma}}_{k,n} \bar{\mathbf{V}}_{k,n}^H$. To realize $\tilde{\mathbf{H}}_{e,k,n} \mathbf{P}_{D,k,n} = \mathbf{0}$, it suffices to set

$$\mathbf{P}_{D,k,1,n} = \bar{\mathbf{V}}_{k,n}[:, 1 : M_u], \text{rank}(\tilde{\mathbf{V}}_{e,k,n}) > M_u. \quad (27)$$

Proof: See [15]. ■

2) *2nd-step digital-domain processing*: With the help of $\mathbf{P}_{D,k,1,n}$, the effective channel between the UE- k and BS at subcarrier- n becomes

$$\hat{\mathbf{H}}_{eff,k,n} = \mathbf{W}_{R,k}^H \mathbf{H}_{k,n} \mathbf{P}_R \mathbf{P}_{D,k,1,n}. \quad (28)$$

To optimize EEMI, the well-known optimal² precoder and combiner are the right and left singular matrices of $\hat{\mathbf{H}}_{eff,k,n}$, denoted as $\hat{\mathbf{U}}_{eff,k,n}$ and $\hat{\mathbf{V}}_{eff,k,n}$, respectively. Lastly, by taking the power constraint into account, the digital precoders are designed as

$$\mathbf{P}_{D,k,n} = \sqrt{M_u} \frac{\mathbf{P}_{D,k,1,n} \hat{\mathbf{V}}_{eff,k,n}}{\|\mathbf{P}_R \mathbf{P}_{D,k,1,n} \hat{\mathbf{V}}_{eff,k,n}\|_F} \quad (29a)$$

$$\mathbf{W}_{D,k,n} = \hat{\mathbf{U}}_{eff,k,n}. \quad (29b)$$

Up to point, we have accomplished the entire HBD design.

²The optimal channel decomposition in terms of the EEMI is derived under the white Gaussian noise. Although $\boldsymbol{\xi}_{k,n} = \mathbf{W}_{R,k} \boldsymbol{\eta}_{k,n}$ might not be strictly white, a good design makes $\mathbf{W}_{R,k}$ approximately semi-unitary in mMIMO setup [22]. In other word, $\boldsymbol{\xi}_{k,n}$ is approximately white.

$$\text{tr}((\mathbf{H}_{e,n}\mathbf{H}_{e,n}^H)^{-1}) = \text{tr}(\mathbf{P}_{n/ab}^{-1}) + \frac{\Re\left\{P_R[a,b](\widehat{\mathbf{H}}_{e,n}^H \mathbf{P}_{n/ab}^{-2} \widehat{\mathbf{H}}_{e,n})[b,:]\tilde{\mathbf{P}}_{R/ab}[:,a]\right\}}{1 + \Re\left\{P_R[a,b](\widehat{\mathbf{H}}_{e,n}^H \mathbf{P}_{n/ab}^{-1} \widehat{\mathbf{H}}_{e,n})[b,:]\tilde{\mathbf{P}}_{R/ab}[:,a]\right\}} \quad (20)$$

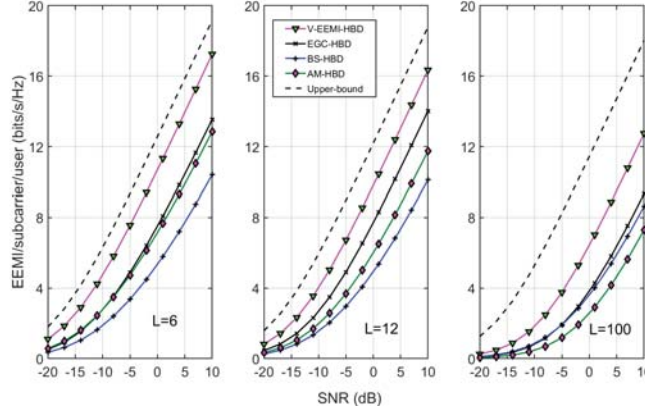


Fig. 2: The comparisons of averaged EEMI among different HBD schemes in frequency-selective channels with $L = 6, 12$ and 100 .

IV. SIMULATIONS

In this section, we provide some numerical results to demonstrate the advantages of the proposed V-EEMI-HBD scheme over existing counterparts. $N_t = 32$ and $N_r = 16$ antennas are employed at the BS and each UE, respectively. The channel related parameters are set as $D = 8$ and $N = 32$, with $h(\cdot)$ being the raised-cosine filter whose roll-off factor $\beta = 0.2$. $M_b = 8$ RF chains are deployed at the BS to serve $K = 4$ UEs, each having $M_u = 2$ RF chains. $\text{Times} = 5$ and $\varepsilon = 0.01$ for EWU algorithm. Without a special specification, 3-bit APS is adopted at the transceivers when applying V-EEMI-HBD.

A. EEMI comparisons in frequency-selective channels

In the part, we compare the achievable EEMI in frequency-selective channels. Three different degrees of channel sparsity are mentioned by setting $L = 6$, $L = 12$, and $L = 100$. For BS-HBD [5], the beam codewords are selected from DFT matrix. For EGC-HBD [14]³ and AM-HBD scheme [17], 5-bit APS is used at the BS and 4-bit APS is used at each UE. From Fig. 2, it can be observed that in sparse channels ($L = 6$ and $L = 12$), the proposed V-EEMI-HBD is remarkably superior to other counterparts, and the performance gap is less than 3dB compared to the ideal upper-bound at high SNR. As L becomes large enough ($L = 100$ in rich scattering environments), all HBD schemes suffer from some EEMI drop due to limitation of hybrid structure. However, V-EEMI-HBD still dramatically outperforms other schemes. Especially when compared with EGC-HBD, a well-recognized excellent HBD solution to non-sparse channels, the advantage is almost 5dB

³For a fair comparison, we have extended original EGC-HBD to the multi-carrier systems, where the common RF precoder is calculated as the quantized Karcher mean of those RF precoders independently designed at each subcarrier.

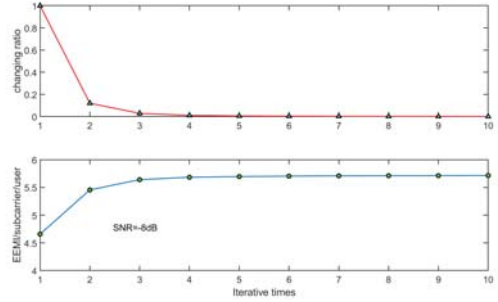


Fig. 3: The relative changing ratio (e in EWU) and averaged EEMI at $\text{SNR} = -8\text{dB}$ versus iterations

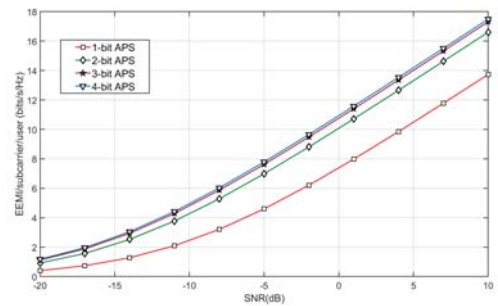


Fig. 4: The comparisons of averaged EEMI versus APS resolution.

at high SNR. Meanwhile, like all the counterparts, V-EEMI-HBD also keeps a linear rather than an exponential complexity *w.r.t* the number of UEs.

B. EEMI comparisons with different resolution of APS

To verify the functionality of EWU RF-domain processing algorithm we fix $L = 6$, then plot the relative changing ratio versus iterations in Fig. 3. The results show that three iterations

suffice to reach a local optimum, implying that the iteration process will not cause a severe computational burden. For SNR=−8dB, it can be observed that three iterations could bring in an over 20% EEMI improvement, demonstrating the effectiveness of the algorithm in enhancing EEMI. In Fig. 4, we further compare the EEMI performance versus the resolution of APS. By observing Fig. 4 together with the leftmost part of Fig. 2, we find that with the proposed HBD, the achievable EEMI in 2-bit-APS case has already been superior to other alternatives. 3-bit APS is basically sufficient to practical use for the excellent EEMI performance and economic hardware cost.

V. CONCLUSION

This work has investigated the transceiver design problem for downlink mmWave multi-user systems. Targeting at maximizing the system EEMI, a high-performance yet low-complexity HBD scheme has been innovatively devised by accounting for the uniqueness of hybrid OFDM systems. In addition to an remarkable advantage over existing alternatives in terms of the EEMI, The entire HBD design does not impose any requirement on either the angular resolution or the channel sparsity. Therefore, the proposed scheme is a promising candidate for downlink multi-user transmission.

VI. ACKNOWLEDGEMENT

This work was in part supported by the Ministry National Key Research and Development Project under Grant 2017YFE0121400, Guandong Key R&D Project under Grant 2018B030338001 and 2019B010153003, and the National Science Foundation under Grant CPS-1932413 and ECCS-1935915.

REFERENCES

- [1] T. Rappaport, J. N. Murdock, and F. Gutierrez, “State of the art in 60-ghz integrated circuits and systems for wireless communications,” *Proc. the IEEE*, vol. 99, no. 8, pp. 1390–1436, August 2011.
- [2] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, January 2013.
- [3] F. Boccardi, R. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.
- [4] X. Gao, L. Dai, S. Han, C.-L. I, and R. Heath, “Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1099, April 2016.
- [5] A. Alkhateeb, G. Leus, and R. Heath, “Limited feedback hybrid precoding for multi-user millimeter wave systems,” *IEEE Trans. Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, November 2015.
- [6] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.
- [7] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Trans. Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, March 2014.
- [8] S. Gao, X. Cheng, and L. Yang, “Spatial multiplexing with limited RF chains: Generalized beamspace modulation (GBM) for mmwave massive MIMO,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 9, pp. 2029–2039, July 2019.
- [9] A. Alkhateeb and R. Heath, “Frequency selective hybrid precoding for limited feedback millimeter wave systems,” *IEEE Trans. Communications*, vol. 64, no. 5, pp. 1801–1818, May 2016.

- [10] Y. Dong, C. Chen, N. Yi, S. Gao, and Y. Jin, “Low-complexity hybrid precoding design for MIMO-OFDM millimeter wave communications,” *IEICE Transactions on Communications*, vol. 100, no. 8, pp. 1228–1237, August 2017.
- [11] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 461–471, February 2004.
- [12] L. Liang, W. Xu, and X. Dong, “Low-complexity hybrid precoding in massive multiuser MIMO systems,” *IEEE Wireless Communication Letters*, vol. 35, no. 5, pp. 653–656, December 2014.
- [13] A. Sayeed and J. Brady, “Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies,” in *Proc. Global Telecommunications Conf.*, Atlanta, GA, December 2013, pp. 3679–3684.
- [14] W. Ni and X. Dong, “Hybrid block diagonalization for massive multiuser MIMO systems,” *IEEE Trans. Communications*, vol. 64, no. 1, pp. 201–211, January 2016.
- [15] R. Rajashekar and L. Hanzo, “Iterative matrix decomposition aided block diagonalization for mm-wave multiuser MIMO systems,” *IEEE Trans. Wireless Communications*, vol. 16, no. 3, pp. 1372–1384, March 2017.
- [16] Q. Zhang, Y. Liu, G. Xie, J. Gao, and K. Liu, “An efficient hybrid diagonalization for multiuser mmwave massive MIMO systems,” in *Proc. of Global Symposium on Millimeter Waves (GSMM)*, Boulder, CO, USA, May 22-24, 2018.
- [17] H. Yuan, J. An, N. Yang, K. Yang, and T. Duong, “Low complexity hybrid precoding for multiuser millimeter wave systems over frequency selective channel,” *VT*, vol. 68, no. 1, pp. 983–987, January 2019.
- [18] B. Liu and H. Zhu, “Rotman lens-based two-tier hybrid beamforming for wideband mmwave MIMO-OFDM system with beam squint,” *EURASIP J. Wireless Commun. Netw.*, vol. 267, pp. 1–13, 2018.
- [19] K. Venugopal, A. Alkhateeb, N. Gonzalez-Prelcic, and R. Heath, “Channel estimation for hybrid architecture-based wideband millimeter wave systems,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1996–2009, September 2017.
- [20] J. Fernandez, N. Prelcic, K. Venugopal, and R. Heath, “Frequency-domain compressive channel estimation for frequency-selective hybrid mmwave MIMO system,” *IEEE Trans. Wireless Communications*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [21] J. C. Chen, “Hybrid beamforming with discrete phase shifters for millimeter-wave massive MIMO systems,” *IEEE Trans. Vehicular Tech.*, vol. 66, no. 8, pp. 7604–7608, August 2017.
- [22] F. Sohrabi and W. Yu, “Hybrid digital and analog beamforming design for large-scale antenna arrays,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, April 2016.

APPENDIX A

Proof: Denote $\mathbf{G}_e = (\mathbf{H}_e \mathbf{H}_e^H)^{-1}$, then I_L can be rewritten as $\sum_{i=1}^{NKM_u} \log_2 \left(1 + \frac{1}{\mathbf{G}_e[i, i] M_u \sigma^2} \right)$. It can be readily verified that $\log_2(1 + 1/x)$ is a convex function over $x > 0$. Via Jensen’s inequality, I_L is lower-bounded by

$$\begin{aligned} I_L &= \sum_{i=1}^{NKM_u} \log_2 \left(1 + \frac{1}{\mathbf{G}_e[i, i] M_u \sigma^2} \right) \\ &\geq NKM_u \log_2 \left(1 + \frac{1}{\frac{M_u \sigma^2}{NKM_u} \sum_{i=1}^{NKM_u} \mathbf{G}_e[i, i]} \right) = I. \end{aligned}$$

On the other hand, leveraging converse of Jensen’s inequality, we have $I_L \leq I + C$, where $C = NKM_u(f(a) + f(b) - 2f(\frac{a+b}{2}))$ as long as $\forall i, \mathbf{G}_e[i, i] M_u \sigma^2 \in [a, b]$. Thus C could be chosen as a constant irrelevant to I .

- If $I + C \leq I_U$, to satisfy $I_L \leq \hat{I} \leq I_U$, we can set $\hat{I} = I + C$.
- If $I + C > I_U$, there must exist $w \in (0, 1)$, such that $\hat{I} = w(I + C) + (1 - w)I = I + wC$ meets the bound requirement.

Thus, it is proper to set $\hat{I}$ as the form given in Proposition 1. ■