This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LWC.2018.2839723, IEEE Wireless Communications Letters

Over-Sampling Codebook-Based Hybrid Minimum Sum-Mean-Square-Error Precoding for Millimeter-Wave 3D-MIMO

Jiening Mao, Zhen Gao, Yongpeng Wu, *Senior Member, IEEE*, and Mohamed-Slim Alouini, *Fellow, IEEE*

***Abstract*—Hybrid precoding design is challenging for millimeter-wave (mmWave) massive MIMO. Most prior hybrid precoding schemes are designed to maximize the sum spectral efficiency (SSE), while seldom investigate the bit-error-rate (BER). Therefore, this letter designs an over-sampling codebook (OSC)-based hybrid minimum sum-mean-square-error (min-SMSE) precoding to optimize the BER. Specifically, given the effective baseband channel consisting of the real channel and analog precoding, we first design the digital precoder/combiner based on min-SMSE criterion to optimize the BER. To further reduce the SMSE between the transmit and receive signals, we propose an OSC-based joint analog precoder/combiner (JAPC) design. Simulation results show that the proposed scheme can achieve the better performance than its conventional counterparts.**

***Index Terms*—Multi-user MIMO, mmWave, hybrid precoding, 3D-MIMO, over-sampling codebook (OSC), massive MIMO.**

I. Introduction

Millimeter-wave (mmWave) massive MIMO with hybrid precoding is a promising key technique for 5G [1], [2]. However, the hybrid precoding design can be challenging due to the non-convex constraints introduced by the analog precoding with modulus constraint [3].

Hybrid analog/digital precoding has been investigated in [3]–[8]. Specifically, a compressive sensing based hybrid precoding and an energy efficient hybrid precoding have been proposed in [3] and [4], respectively, while they are limited to single-user MIMO systems. A phased zero forcing hybrid precoding has been proposed for multi-user mmWave MIMO [5], while only single-antenna user is considered. Moreover, to support multi-user mmWave MIMO with each user equipped with multiple antennas and single RF chain, a two-stage multi-user hybrid precoding and a heuristic hybrid precoding have been proposed in [6] and [7], respectively. However, how to effectively support multi-user and multi-stream transmission for each user is not well addressed. Recently, a hybrid block diagonalization (BD) precoding is developed from the traditional BD precoding in [8], which can be used to support multi-user massive MIMO with each user equipped with multiple antennas and multiple RF chains. On the other hand, most prior work [3]–[8] have designed the analog precoding to maximize the sum spectral efficiency (SSE) other than to optimize the bit-error-rate (BER) performance, and they usually consider the ideal uniform linear array (ULA) other than the practical uniform planar array (UPA).

In this letter, we propose an over-sampling codebook (OSC)-based hybrid minimum sum-mean-square-error (min-SMSE) precoding scheme for mmWave multi-user three-dimensional (3D)-MIMO systems to optimize the BER. Specifically, given the effective baseband channel consisting of the real channel and analog precoding, we first design the digital precoder/combiner to minimize the SMSE between the transmit symbols and receive symbols. Furthermore, we propose an OSC-based joint analog precoder/combiner (JAPC) design by further reducing the SMSE for improved BER. Simulation results confirm that the proposed scheme has the better performance than the conventional schemes. To the best of our knowledge, this is the first work to jointly design digital and analog precoder/combiner for mmWave massive multi-user MIMO based on minimum SMSE criterion.

Notations: Lower-case and upper-case boldface letters denote vectors and matrices, respectively; $(\cdot)^T$, $(\cdot)^H$, and $\mathrm{tr}(\cdot)$ denote the transpose, conjugate transpose, and trace of a matrix, respectively; $(\cdot)^{(i)}$ and $(\cdot)_{i,j}$ represent the i-th column and i-th row and j-th column element of a matrix, respectively; $\mathbf{I}_N$ represents the $N\times N$ identity matrix; $\|\cdot\|_2$, $\|\cdot\|_{\mathrm{F}}$, $\mathrm{diag}(\cdot)$, $\mathrm{angle}(\cdot)$, and $\mathbb{E}[\cdot]$ represent the ℓ_2-nrom, Frobenius norm, diagonalization, phase extraction, and expectation operators, respectively.

II. System Model

We consider the downlink mmWave multi-user 3D-MIMO system with the hybrid precoding. Assume that the base station (BS) employs N_t antennas but $M_t \ll N_t$ RF chains to simultaneously serve K users, and each user employs N_r antennas but $M_r \ll N_r$ RF chains to support $N_s \le M_r$ data streams. Therefore, the BS can simultaneously support $KN_s \le M_t$ data streams. In the downlink transmission with $KM_r = M_t$, the received signal at the k-th user can is

$$\mathbf{y}_k = \mathbf{V}_k^H\mathbf{M}_k^H\left(\gamma\mathbf{H}_k\mathbf{F}\mathbf{W}\mathbf{x}+\mathbf{n}_k\right), \quad (1)$$

where $\mathbf{V}_k \in \mathbb{C}^{M_r\times N_s}$ and $\mathbf{M}_k \in \mathbb{C}^{N_r\times M_r}$ represent the digital combiner and analog combiner of the k-th user, respectively, $\mathbf{H}_k \in \mathbb{C}^{N_r\times N_t}$ is the downlink channel associated with the BS and the k-th user, $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_K] \in \mathbb{C}^{N_t\times M_t}$ and $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_K] \in \mathbb{C}^{M_t\times KN_s}$ represent the analog precoder and digital precoder at the BS, respectively, $\mathbf{F}_k \in \mathbb{C}^{N_t\times M_r}$ and $\mathbf{W}_k \in \mathbb{C}^{M_t\times N_s}$ $(1\le k\le K)$ represent the analog precoder and digital precoder for the k-th user, respectively. γ is the normalization factor so that $\gamma^2\mathrm{tr}\left(\mathbf{F}\mathbf{W}\mathbf{W}^H\mathbf{F}^H\right) = P_t$, and P_t is the transmit power. $\mathbf{x} = \left[\mathbf{x}_1{}^T, \mathbf{x}_2{}^T, \cdots, \mathbf{x}_K{}^T\right]^T \in \mathbb{C}^{KN_s\times 1}$ represents the transmitted signal vector for K users. $\mathbf{n}_k \in \mathbb{C}^{N_r\times 1}$ is the additive white Gaussian noise (AWGN) vector at the k-th user.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61701027 and 61701301), the Beijing Natural Science Foundation (Grant No. 4182055), and Huawei Innovation Research Program (HIRP). (Corresponding author: Zhen Gao)

J. Mao and Z. Gao are with School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (E-mails: 2220170402@bit.edu.cn, gaozhen16@bit.edu.cn).

Y. Wu is with Department aof Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (E-mail: yongpeng.wu2016@gmail.com).

M. -S. Alouini is with King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia (E-mail: slim.alouini@kaust.edu.sa).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LWC.2018.2839723, IEEE Wireless Communications Letters

The mmWave MIMO channel matrix $\mathbf{H}$ is assumed to be a sum of the contributions of N_c scattering clusters, and each cluster contains N_p propagation paths [3], i.e.,

$$\mathbf{H}_k = \sqrt{\frac{N_t N_r}{N_c N_p}} \sum_{i=1}^{N_c} \sum_{l=1}^{N_p} \alpha_{il}^k \mathbf{a}_r^k \left(\theta_{il}^{rk}, \phi_{il}^{rk}\right) \mathbf{a}_t^k \left(\theta_{il}^{tk}, \phi_{il}^{tk}\right)^H, \quad (2)$$

where α_{il}^k is the complex gain of the i-th path in the l-th cluster and $\alpha_{il}^k \sim \mathcal{CN}(0,1)$. $\theta_{il}^{rk}\left(\phi_{il}^{rk}\right)$ and $\theta_{il}^{tk}\left(\phi_{il}^{tk}\right)$ are the azimuth (elevation) angles of arrival (AoAs) and departure (AoDs) of the i-th path in the l-th cluster, respectively. The angles in each cluster follow the uniform distribution and have the constant angle spreads (standard deviation), which can be denoted by σ_ϕ^{tk}, σ_θ^{tk}, σ_ϕ^{rk}, and σ_θ^{rk}, respectively. $\mathbf{a}_r^k\left(\theta_{il}^{rk}, \phi_{il}^{rk}\right)$ and $\mathbf{a}_t^k\left(\theta_{il}^{tk}, \phi_{il}^{tk}\right)$ are the normalized receive and transmit array response vectors at an azimuth (elevation) angle $\theta_{il}^{rk}\left(\phi_{il}^{rk}\right)$ and $\theta_{il}^{tk}\left(\phi_{il}^{tk}\right)$, respectively. For UPA in the yz-plane with N_y and N_z elements on the y and z axes, respectively, the array response vector can be written as

$$\mathbf{a}_{\text{UPA}}(\theta,\phi) = \frac{1}{\sqrt{N_y N_z}} \Big[1, \cdots, e^{j\frac{2\pi}{\lambda} d(n \sin(\theta)\cos(\phi) + m \sin(\phi))}, \cdots, e^{j\frac{2\pi}{\lambda} d((N_y-1)\sin(\theta)\cos(\phi) + (N_z-1)\sin(\phi))}\Big]^T, \quad (3)$$

where $1 \le n < N_y$ and $1 \le m < N_z$ are the y and z indices of an antenna element, respectively. Finally, λ and d denote the wavelength and adjacent antenna spacing, respectively.

III. OSC-Based Hybrid Min-SMSE Precoding

In this section, we first propose a min-SMSE criterion based digital precoding design. To further improve the BER performance by reducing SMSE, we propose a JAPC design based on our proposed OSC in the analog part.

A. Digital Precoding Design

Assume that the k-th user's effective baseband channel $\mathbf{H}_{\text{eff}}^k = \mathbf{M}_k^H \mathbf{H}_k \mathbf{F} \in \mathbb{C}^{M_r \times M_t}$ $(1 \le k \le K)$ is given, the estimated signal at the k-th user can be rewritten as

$$\hat{\mathbf{x}}_k = \mathbf{y}_k/\gamma = \mathbf{V}_k^H \mathbf{H}_{\text{eff}}^k \mathbf{W}\mathbf{x} + \mathbf{V}_k^H \mathbf{M}_k^H \mathbf{n}_k/\gamma. \quad (4)$$

Then the k-th user's the mean square error (MSE) can be expressed as $\xi_k = \mathbb{E}\left[\|\hat{\mathbf{x}}_k - \mathbf{x}_k\|_2^2\right]$. Moreover, since $\mathbf{x}_k$ and $\mathbf{n}_k$ are mutually independent, ξ_k can be further expressed as

$$\xi_k = \text{tr}\Big(\mathbf{V}_k^H \mathbf{H}_{\text{eff}}^k \mathbf{W}\mathbf{W}^H \big(\mathbf{H}_{\text{eff}}^k\big)^H \mathbf{V}_k\Big) + \frac{\sigma^2}{\gamma^2}\text{tr}\big(\mathbf{V}_k^H \mathbf{M}_k^H \mathbf{M}_k \mathbf{V}_k\big) - \text{tr}\left(\mathbf{V}_k^H \mathbf{H}_{\text{eff}}^k \mathbf{W}_k\right) - \text{tr}\left(\mathbf{W}_k^H \left(\mathbf{H}_{\text{eff}}^k\right)^H \mathbf{V}_k\right) + N_s. \quad (5)$$

In this way, the digital precoder/combiner design can be formulated as the following optimization problem

$$\min_{\mathbf{W},\mathbf{V}^H} \sum\nolimits_{k=1}^{K} \xi_k \quad \text{s.t.} \quad \gamma^2 \text{tr}\left(\mathbf{F}\mathbf{W}\mathbf{W}^H\mathbf{F}^H\right) = P_t, \quad (6)$$

where $\mathbf{V} = \text{diag}\left(\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_K\right)$ represents the digital combiner for K users. Moreover, since K users will estimate their respective signals independently in the downlink, we can design the digital combiner for each user separately. Hence, the problem (6) can be decomposed as the subproblems $\mathbf{V}_k^H = \arg\min_{\mathbf{V}_k^H} \xi_k$ for $k = 1, 2, \cdots, K$. Consider the partial derivative of $\mathbf{V}_k^H$ with respect to ξ_k and let it equal 0, we arrive at

$$\mathbf{V}_k^H = \mathbf{W}_k^H \left(\mathbf{H}_{\text{eff}}^k\right)^H \left(\mathbf{H}_{\text{eff}}^k \mathbf{W}\mathbf{W}^H \left(\mathbf{H}_{\text{eff}}^k\right)^H + \frac{\sigma^2}{\gamma^2}\mathbf{M}_k^H \mathbf{M}_k\right)^{-1}. \quad (7)$$

On the other hand, for the digital precoder $\mathbf{W}$ at the BS, the SMSE of K users will be considered, i.e., $\xi = \sum_{k=1}^K \xi_k = \mathbb{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2\right]$. Similar to (5), ξ can be expressed as

$$\xi = \text{tr}\left(\mathbf{V}^H \mathbf{H}_{\text{eff}} \mathbf{W}\mathbf{W}^H \mathbf{H}_{\text{eff}}^H \mathbf{V}\right) - \text{tr}\left(\mathbf{V}^H \mathbf{H}_{\text{eff}} \mathbf{W}\right) - \text{tr}\left(\mathbf{W}^H \mathbf{H}_{\text{eff}}^H \mathbf{V}\right) + \frac{\sigma^2}{\gamma^2}\text{tr}\left(\mathbf{V}^H \mathbf{M}^H \mathbf{M}\mathbf{V}\right) + KN_s. \quad (8)$$

Consider the partial derivative of $\mathbf{W}$ with respect to ξ and let it equal 0, the optimal $\mathbf{W}$ can be

$$\mathbf{W} = \left(\mathbf{H}_{\text{eff}}^H \mathbf{V}\mathbf{V}^H \mathbf{H}_{\text{eff}}\right)^{-1} \mathbf{H}_{\text{eff}}^H \mathbf{V}. \quad (9)$$

From (7) and (9), we can observe that the solutions of $\mathbf{V}_k$ and $\mathbf{W}$ depend on each other. To decouple these two solutions, we first obtain an initial solution of V, denoted by $\mathbf{V}_{\text{ini}} = \text{diag}\left(\mathbf{V}_{\text{ini}}^1, \mathbf{V}_{\text{ini}}^2, \cdots, \mathbf{V}_{\text{ini}}^K\right)$, where each $\mathbf{V}_{\text{ini}}^k \in \mathbb{C}^{N_s \times N_s}$ $(1 \le k \le K)$ can be an arbitrary unitary matrix. Moreover, we can obtain $\mathbf{W}$ by substituting the initial combiner $\mathbf{V}_{\text{ini}}$ into (9), and then obtain $\mathbf{V}_k$ according to (7). Note that different $\mathbf{V}_{\text{ini}}$'s can lead to different digital precoder/combiner solutions $\{\mathbf{W}, \mathbf{V}\}$, but these different solutions $\{\mathbf{W}, \mathbf{V}\}$ are equivalent. Specifically, we first introduce the **Lemma 1** as follows.

Lemma 1: Assume that the matrix $\mathbf{A} = \left[\mathbf{A}_1^T, \cdots, \mathbf{A}_N^T\right]^T$ is invertible. Then

$$\mathbf{A}_i \left(\mathbf{A}^H \mathbf{A}\right)^{-1} \mathbf{A}_j^H = \delta(i-j)\,\mathbf{I}, 1 \le i \le N, 1 \le j \le N. \quad (10)$$

Proof: Due to the invertibility of $\mathbf{A}$, we can obtain $\mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H = \mathbf{A}\mathbf{A}^{-1}\left(\mathbf{A}^H\right)^{-1}\mathbf{A}^H = \mathbf{I}$. Meanwhile,

$$\mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H = \text{diag}\Big\{\mathbf{A}_1\left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}_1^H, \cdots, \mathbf{A}_N\left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}_N^H\Big\}. \quad (11)$$

Hence, (10) is proven, where $\delta(i-j) = 1$ only if $j = i$, otherwise $\delta(i-j) = 0$. ■

Given $\{\mathbf{V}_{\text{ini}}^k\}_{k=1}^K$ being arbitrary unitary matrices, $\mathbf{W} = \left(\mathbf{H}_{\text{eff}}^H \mathbf{H}_{\text{eff}}\right)^{-1} \mathbf{H}_{\text{eff}}^H \mathbf{V}_{\text{ini}}$ and $\mathbf{W}_k = \left(\mathbf{H}_{\text{eff}}^H \mathbf{H}_{\text{eff}}\right)^{-1} \left(\mathbf{H}_{\text{eff}}^k\right)^H \mathbf{V}_{\text{ini}}^k$ according to (9). Then we have $\mathbf{W}\mathbf{W}^H = \left(\mathbf{H}_{\text{eff}}^H \mathbf{H}_{\text{eff}}\right)^{-1}$. Consider the invertibility of matrix $\mathbf{H}_{\text{eff}}$ and **Lemma 1**,

$$\mathbf{V}_k^H \overset{a}{\approx} \mathbf{W}_k^H \left(\mathbf{H}_{\text{eff}}^k\right)^H \left(\mathbf{H}_{\text{eff}}^k \mathbf{W}\mathbf{W}^H \left(\mathbf{H}_{\text{eff}}^k\right)^H + \frac{\sigma^2}{\gamma^2}\mathbf{I}\right)^{-1} = \frac{\gamma^2}{\gamma^2 + \sigma^2} \mathbf{W}_k^H \left(\mathbf{H}_{\text{eff}}^k\right)^H = \frac{\gamma^2}{\gamma^2 + \sigma^2} \left(\mathbf{V}_{\text{ini}}^k\right)^H, \quad (12)$$

where the approximation (a) is due to the asymptotic orthogonality $\mathbf{M}_k^H \mathbf{M}_k \approx \mathbf{I}$ when N_r is large [3]. Given $\mu = \frac{\gamma^2}{\gamma^2 + \sigma^2}$ and $\forall k$,

$$\mathbf{V}_k^H \mathbf{H}_{\text{eff}}^k \mathbf{W}_k \approx \mu \left(\mathbf{V}_{\text{ini}}^k\right)^H \mathbf{H}_{\text{eff}}^k \left(\mathbf{H}_{\text{eff}}^H \mathbf{H}_{\text{eff}}\right)^{-1} \left(\mathbf{H}_{\text{eff}}^k\right)^H \mathbf{V}_{\text{ini}}^k = \mu \left(\mathbf{V}_{\text{ini}}^k\right)^H \mathbf{V}_{\text{ini}}^k = \mu \mathbf{I}_{N_s}. \quad (13)$$

Similarly, given $1 \le m \le K$, $1 \le k \le K$, and $m \ne k$, $\mathbf{V}_m^H \mathbf{H}_{\text{eff}}^m \mathbf{W}_k \approx \mathbf{0}$. Hence we can have $\mathbf{V}^H \mathbf{H}_{\text{eff}} \mathbf{W} \approx \mu \mathbf{I}_{KN_s}$, which does not depend on $\{\mathbf{V}_{\text{ini}}^k\}_{k=1}^K$ if $\{\mathbf{V}_{\text{ini}}^k\}_{k=1}^K$ are arbitrary unitary matrices.

B. Analog Precoding Design

By substituting $\mathbf{V}^H \mathbf{H}_{\text{eff}} \mathbf{W} \approx \mu \mathbf{I}_{KN_s}$ into (8), the SMSE ξ can be further expressed as

$$\begin{aligned}\xi &\approx \left(\mu^2 - 2\mu + 1\right) KN_s + \frac{\sigma^2}{\gamma^2}\text{tr}\left(\mathbf{V}^H \mathbf{M}^H \mathbf{M}\mathbf{V}\right) \\ &\overset{b}{=} \left(\mu^2 - 2\mu + 1\right) KN_s + \frac{\sigma^2}{P_t}\text{tr}\left(\mathbf{F}\mathbf{W}\mathbf{W}^H\mathbf{F}^H\right)\text{tr}\left(\mathbf{V}^H \mathbf{M}^H \mathbf{M}\mathbf{V}\right) \\ &\overset{c}{\approx} \left(\mu^2 - 2\mu + 1\right) KN_s + \frac{KN_s \mu^2 \sigma^2}{P_t}\text{tr}\left(\left(\mathbf{H}_{\text{eff}}^H \mathbf{H}_{\text{eff}}\right)^{-1}\right), \quad (14)\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LWC.2018.2839723, IEEE Wireless Communications Letters

where the equation (b) is due to the power constraint in (6) and the approximation (c) is due to $\mathbf{V}_k\mathbf{V}_k^H \approx \mu^2\mathbf{I}_{N_s}$ and $\mathbf{V}\mathbf{V}^H \approx \mu^2\mathbf{I}_{KN_s}$ based on (12). Note that $\mathbf{F}^H\mathbf{F} \approx \mathbf{I}$ ($\mathbf{M}^H\mathbf{M} \approx \mathbf{I}$) when N_t (N_r) is large [3].

In (14), when signal-to-noise ratio (SNR) is high with $\sigma^2 \to 0$, we can obtain $\mu \to 1$ and $\xi \approx \frac{KN_s\sigma^2}{P_t}\mathrm{tr}\left(\left(\mathbf{H}_{\text{eff}}^H\mathbf{H}_{\text{eff}}\right)^{-1}\right)$. Note that when SNR is low where the noise dominates the performance, i.e., $\sigma^2 \to \infty$, and then we have $\mu \to 0$ and $\xi \approx KN_S$, which is a constant. Hence, we consider the analog precoding design at high SNR. Moreover, we consider the singular value decomposition (SVD) of $\mathbf{H}_{\text{eff}}$, i.e., $\mathbf{H}_{\text{eff}} = \mathbf{U}\mathbf{\Sigma}\mathbf{D}^H$, where $\mathbf{\Sigma} = \mathrm{diag}\left\{\sigma_1, \sigma_2, \cdots, \sigma_{KN_s}\right\}$ with the descent order, and then ξ can be rewritten as

$$\xi \approx \frac{KN_s\sigma^2}{P_t}\mathrm{tr}\left(\mathbf{\Sigma}^{-2}\right) = \frac{KN_s\sigma^2}{P_t}\sum\nolimits_{i=1}^{KN_s}\sigma_i^{-2}. \tag{15}$$

To further optimize the SMSE, the analog precoding design can be formulated as the following optimization problem

$$\begin{aligned}&\min_{\{\mathbf{M}_k\}_{k=1}^K,\mathbf{F}} \sum\nolimits_{i=1}^{KN_s}\sigma_i^{-2}\\ &\text{s.t. angle}\left(\left(\mathbf{M}_k\right)_{i,j}\right) \in \mathcal{Q}_r, \text{angle}\left(\mathbf{F}_{i,j}\right) \in \mathcal{Q}_t, \forall i,j,k,\end{aligned} \tag{16}$$

where $\mathcal{Q}_r$ and $\mathcal{Q}_t$ are the candidate sets of quantized phases due to the finite resolution of the phase shifters. From (16), we observe that given $\sum_{i=1}^{KN_s}\sigma_i^{-2}$, $\mathbf{H}_{\text{eff}}$ with equal singular values will reach the minimization of (16). This enlightens us to design the analog precoder/combiner with the condition number $\mathrm{cond}\left(\mathbf{H}_{\text{eff}}\right)$ as small as possible, where $\mathrm{cond}\left(\mathbf{H}_{\text{eff}}\right) = \sigma_1/\sigma_{KN_s}$. On the other hand, given the equal singular values, the minimization of (16) is equivalent to

$$\begin{aligned}&\max_{\{\mathbf{M}_k\}_{k=1}^K,\mathbf{F}} \sum\nolimits_{i=1}^{KN_s}\sigma_i^{2}\\ &\text{s.t. angle}\left(\left(\mathbf{M}_k\right)_{i,j}\right) \in \mathcal{Q}_r, \text{angle}\left(\mathbf{F}_{i,j}\right) \in \mathcal{Q}_t, \forall i,j,k.\end{aligned} \tag{17}$$

Moreover, even the optimization problem (17) can also be challenging, which requires the prohibitively high complexity with $\mathcal{O}\left(2^{KB_rM_rN_rB_tM_tN_t}\right)$, where B_r and B_t are the quantization bits of phase shifters at the users and BS, respectively. While if $\mathbf{H}_{\text{eff}}$ is a diagonally-dominant matrix, the optimization problem (17) can be approximated to $\max_{\{\mathbf{M}_k\}_{k=1}^K,\mathbf{F}} \sum_{i=1}^{KN_s}\left|\left(\mathbf{H}_{\text{eff}}\right)_{i,i}\right|^2$. Based on the analysis above, we consider the following multi-object optimization problem

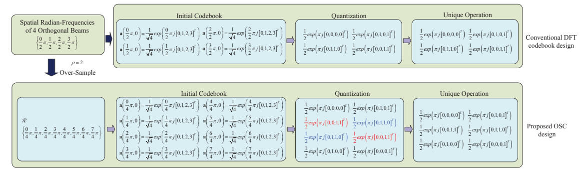

Fig. 1. An example of the proposed OSC design, where a 4-element ULA with $\rho = 2$ and $q = 1$ bit is considered.

$$\begin{aligned}&\max_{\{\mathbf{M}_k\}_{k=1}^K,\mathbf{F}} \mathbf{g}\left(\{\mathbf{M}_k\}_{k=1}^K, \mathbf{F}\right) = \left[\sum_{i=1}^{KN_s}\left|\left(\mathbf{H}_{\text{eff}}\right)_{i,i}\right|^2, -\mathrm{cond}\left(\mathbf{H}_{\text{eff}}\right)\right]^T\\ &\text{s.t. } \mathbf{M}_k^{(m)} \in \mathcal{D}_{k,r}, \mathbf{F}^{(m)} \in \mathcal{D}_t, \forall m, k,\end{aligned} \tag{18}$$

where $\mathcal{D}_{k,r}$ and $\mathcal{D}_t$ are the predefined codebooks for $\mathbf{M}_k$ and $\mathbf{F}$, respectively. Thus, the computational complexity to acquire the solution to (18) can be reduced to $\mathcal{O}\left(K|\mathcal{D}_{k,r}|_c|\mathcal{D}_t|_c\right)$, and $|\cdot|_c$ denotes the cardinality of a set. To ensure $\mathbf{H}_{\text{eff}}$ to be a diagonally-dominant matrix, we propose that each candidate from $\mathcal{D}_{k,r}$ and $\mathcal{D}_t$ has the form of UPA response vector as shown in (3), i.e.,

$$\mathbf{a}\left(w_y, w_z\right) = \frac{1}{\sqrt{N}}\left[1, \cdots, e^{j(nw_y+mw_z)}, \cdots, e^{j((N_y-1)w_y+(N_z-1)w_z)}\right]^T, \tag{19}$$

where $N_yN_z = N$, w_y and w_z denote the spatial frequencies associated with the azimuth and elevation angles, respectively.

The conventional 2D-DFT codebook has the limited spatial resolution. To further improve the spatial resolution, we design an OSC as shown in **Algorithm 1**, where the set $\mathrm{unique}(\mathcal{D})$ has the same candidates as the set $\mathcal{D}$ without repetitions. Specifically, we first obtain an OSC $\mathcal{D}$ given the over-sampling factor ρ (steps 1∼2). For each candidate in $\mathcal{D}$, we quantize the phases of all elements by considering the quantization bits of phase shifters q (step 3), and finally make candidates in $\mathcal{D}$ be unique (step 4). Conventional 2D-DFT codebook is a special case of the proposed OSC for UPA when $\rho = 1$. When ρ becomes large, the proposed codebook can achieve the better spatial resolution as illustrated in Fig. 1. On the other hand, the better spatial resolution indicates the stronger

---

Algorithm 1 Proposed OSC Design

Input: Number of antennas $N = N_yN_z$, quantization bits of phase shifters q, and the over-sampling factor ρ.

Output: Predefined codebook $\mathcal{D}$.

1. Obtain the candidate set of practical quantized phases $\mathcal{Q} = \left\{0, \frac{2\pi}{2^q}, \cdots, \frac{2\pi(2^q-1)}{2^q}\right\}$ and the sets of over-sampling spatial radian-frequencies associated with the azimuth and elevation angles $\mathcal{R}_y = \left\{0, \frac{2\pi}{\rho N_y}, \cdots, \frac{2\pi(\rho N_y-1)}{\rho N_y}\right\}$, $\mathcal{R}_z = \left\{0, \frac{2\pi}{\rho N_z}, \cdots, \frac{2\pi(\rho N_z-1)}{\rho N_z}\right\}$.
2. Obtain the codebook $\mathcal{D}=\left\{\mathbf{a}\left(w_y, w_z\right) | w_y \in \mathcal{R}_y, w_z \in \mathcal{R}_z\right\}$.
3. Quantize phases of all candidates in $\mathcal{D}$ as $\text{angle}(\mathbf{a}(w_y,w_z)_{i,j}) = \arg\min_{\alpha\in\mathcal{Q}}\left\|\text{angle}(\mathbf{a}(w_y,w_z)_{i,j}) - \alpha\right\|_2^2$, where $\mathbf{a}\left(w_y, w_z\right)$ is one specific candidate in $\mathcal{D}$, $1 \le i \le N$, and $j = 1$.
4. $\mathcal{D} = \mathrm{unique}(\mathcal{D})$ to make candidates in $\mathcal{D}$ be unique.

Algorithm 2 Proposed JAPC Design

Input: Maximum correlation factor β, BS codebook $\mathcal{D}_t$, user codebook $\mathcal{D}_{k,r}$, and channel $\mathbf{H}_k$ for $1 \le k \le K$.

Output: Analog precoder $\mathbf{F}$ and analog combiner $\mathbf{M}_k$, $\forall k$.

1. Initialization: $\mathcal{K} = \{1, 2, \cdots, K\}$, $\mathbf{F} = \mathbf{M}_k = \Phi, \forall k$, and Φ denotes empty matrix;
2. **for** $i_{\text{iter}} = 1 : KM_r$
3. $\{k^*, \mathbf{a}_r^*, \mathbf{a}_t^*\} = \arg\max_{k\in\mathcal{K},\mathbf{a}_r\in\mathcal{D}_{k,r},\mathbf{a}_t\in\mathcal{D}_t}\left\|\mathbf{a}_r^H\mathbf{H}_k\mathbf{a}_t\right\|_2^2$;
4. $\mathbf{M}_{k^*} = [\mathbf{M}_{k^*}|\mathbf{a}_r^*]$, $\mathbf{F}_{k^*} = [\mathbf{F}_{k^*}|\mathbf{a}_t^*]$;
5. $\mathcal{D}_{k^*,r} = \mathrm{setdiff}\left(\mathcal{D}_{k^*,r}, \left\{\mathbf{a}_r | \mathbf{a}_r \in \mathcal{D}_{k^*,r}, \left|(\mathbf{a}_r^*)^H\mathbf{a}_r\right| \ge \beta\right\}\right)$;
6. $\mathcal{D}_t = \mathrm{setdiff}\left(\mathcal{D}_t, \left\{\mathbf{a}_t | \mathbf{a}_t \in \mathcal{D}_t, \left|(\mathbf{a}_t^*)^H\mathbf{a}_t\right| \ge \beta\right\}\right)$;
7. $\mathcal{K} = \mathrm{setdiff}\left(\mathcal{K}, \{k^*\}\right)$ when the number of columns of $\mathbf{M}_{k^*}$ equals M_r;
8. **end for**
9. $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_K]$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LWC.2018.2839723, IEEE Wireless Communications Letters

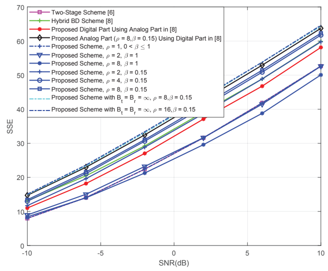

Fig. 2. SSE achieved by different schemes in a multi-user MIMO system, where $K = 8$, $M_t = 8$, $M_r = N_s = 1$.

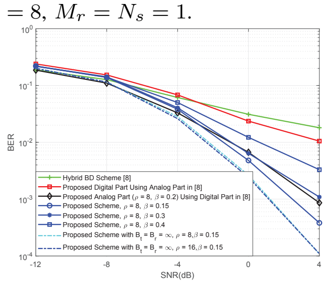

Fig. 3. BER achieved by different schemes in a multi-user MIMO system, where $K = 2$, $M_t = 4$, $M_r = N_s = 2$.

correlation among candidates in $\mathcal{D}$. This implies that directly acquiring analog precoder/combiner $\mathbf{M}_k$ and $\mathbf{F}_k$ to maximize $\sum_{i=1}^{KN_s} \left|(\mathbf{H}_{\text{eff}})_{i,i}\right|^2$ will lead $\mathbf{H}_{\text{eff}}^k$ to be ill-conditioned or even rank-insufficient.

To solve the optimization problem (18), we further propose a heuristic JAPC algorithm as shown in **Algorithm 2**, where the set $\text{setdiff}(\mathcal{A}, \mathcal{B})$ has all candidates in $\mathcal{A}$ that are not in $\mathcal{B}$. Specifically, in each iteration, we first acquire the user index k^* and the associated $\{\mathbf{a}_r^*, \ \mathbf{a}_t^*\}$ that maximize the diagonal elements of effective channels (step 3) to be part of the analog precoder/combiner matrices $\mathbf{M}_k$ and $\mathbf{F}_k$, respectively (step 4). To prevent $\mathbf{H}_{\text{eff}}$ to be ill-conditioned, the candidates in $\mathcal{D}_{k,r}$ and $\mathcal{D}_t$ having strong correlation with the selected $\{\mathbf{a}_r^*, \ \mathbf{a}_t^*\}$ in **step 3** are removed (steps 5$\sim$6). Finally, if the k^*-th user's analog combiner matrix $\mathbf{M}_k^*$ is completed, the index k^* is removed from $\mathcal{K}$ (step 7).

IV. Simulation Results

The simulation parameters are shown as follows: BS and users are equipped with 8×8 and 4×4 UPAs, respectively, $B_t = 3$ bits, $B_r = 2$ bits, and we consider the 16-QAM modulation. For the mmWave channel model, we consider $N_c = 8$ and $N_p = 10$, azimuth/elevation AoAs and AoDs follow the uniform distribution $\mathcal{U}\left[-\pi/2, \pi/2\right]$, and all the angle spreads are 7.5° [3], [8]. We will investigate the performance of the proposed scheme by comparing it with the two-stage hybrid precoding scheme [6], the hybrid BD scheme [8], the proposed digital part using analog part in [8], and our proposed analog part using digital part (i.e., BD precoding) in [8]. Besides, the proposed scheme with $B_t = B_r = \infty$ (i.e., the step 3 and step 4 in Algorithm 1 are removed) are simulated to investigate the performance gap compared to the optimal schemes.

Fig. 2 and Fig. 3 investigate the SSE and BER performance, respectively. In Fig. 2, the proposed scheme with $\rho = 8$ and $\beta = 0.15$ has the better SSE than the two-stage scheme [6] and the hybrid BD scheme [8]. Especially, the proposed OSC-based JAPC design with the digital part in [8] outperforms the hybrid BD scheme [8]. Besides, the SSE of the proposed scheme with $\beta = 1$ becomes worse when ρ increases, since directly using solution to (18) with OSC will lead large $\text{cond}\left(\mathbf{H}_{\text{eff}}\right)$. While the SSE of the proposed scheme with $\beta = 0.15$ becomes better when ρ increases, since larger ρ can achieve better spatial resolution and the proposed JAPC design with a suitable β can guarantee small $\text{cond}\left(\mathbf{H}_{\text{eff}}\right)$. Note that when $\rho = 1$, the OSC is reduced to conventional 2D-DFT codebook, and the JAPC design with $\rho = 1$ for $0 < \beta \leq 1$ is equivalent since elements in 2D-DFT codebook are orthogonal. In Fig. 3, we observe that when BER $= 10^{-2}$ is considered, the proposed scheme ($\rho = 8$ and $\beta = 0.15$), the proposed min-SMSE based digital part using analog part in [8], and the proposed OSC-based JAPC design using analog part in [8] have over 5.1 dB, 1 dB, and 5 dB SNR gains than the hybrid BD scheme, respectively. Note that when $B_t = B_r = \infty$, the proposed scheme with $\rho = 8$ and the proposed scheme with $\rho = 16$ have very similar BER and SSE performance, which indicates OSC with $\rho = 8/16$ can have sufficiently large spatial resolution and these two curves can be considered as the optimal performance with infinite spatial resolution of codebook and infinite quantization of phase shifters. The BER and SSE performance of the proposed scheme with $\rho = 8$ and $\beta = 0.15$ still has a gap compared to the optimal schemes due to the limited quantization of phase shifters.

V. Conclusions

This letter has proposed an OSC-based hybrid min-SMSE precoding scheme to support multi-user and multi-stream transmission for each user in mmWave 3D-MIMO systems. Specifically, we design the digital precoder/combiner to improve the BER performance based on the min-SMSE criterion. Moreover, in the analog part, we propose an OSC and the associated JAPC design by further reducing the SMSE. Finally, simulation results show the proposed scheme can achieve the better performance than the conventional schemes.

References

[1] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, M. Z. Shakir, "MmWave massive MIMO based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.* vol. 22, no. 5, pp. 13-21, Oct. 2015.

[2] Y. Sun and C. Qi, "Weighted sum-rate maximization for analog beamforming and combining in millimeter wave massive MIMO communications," *IEEE Wireless Commun. Lett.* vol. 21, no. 8, pp. 1883-1886, Oct. 2017.

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

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

[5] 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, Oct. 2014.

[6] 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.

[7] F. Sohrabi, 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.

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