Multilevel-DFT based Low-Complexity Hybrid Precoding for Millimeter Wave MIMO Systems

Yu-Hsin Liu, Chiang-Hen Chen, Cheng-Rung Tsai, and An-Yeu (Andy) Wu, Fellow, IEEE
Graduate Institute of Electronics Engineering
National Taiwan University
Taipei, Taiwan
{mike, james, raulshepherd, andywu}@access.ee.ntu.edu.tw

Abstract—The millimeter wave (mmWave) multiple-input multiple-output (MIMO) is a promising technology for next-generation (5G) communication systems. The inherently high path loss in mmWave can be compensated by adopting large antenna array. However, huge cost of radio frequency (RF) chains at high frequencies, as well as large number of antennas, makes full-digital precoding in mmWave-MIMO system infeasible. Hence, hybrid analog/digital precoding techniques are proposed to reduce the hardware cost of RF components while achieving similar performance as full-digital precoder. Usually it requires iterative process and matrix inversion to split full-digital precoder into hybrid structure. In this paper, we propose a low-complexity hybrid precoding algorithm with predefined rotated multilevel-DFT codebook. It can leverage orthogonal mapping process to avoid iterative process and matrix inversion. Simulation results demonstrate that the proposed algorithm can achieve 97.6% spectrum efficiency compared with state-of-the-art low-complexity hybrid precoding algorithm while reducing 99.7% complexity in terms of complex multiplications.

Keywords—millimeter wave MIMO, hybrid precoding, beamforming, low-complexity.

I. INTRODUCTION

The millimeter wave (mmWave) multiple-input multiple-output (MIMO) is a promising technology for 5G networks [1]. Conventional MIMO precoding is realized at baseband (BB) through full-digital precoder. For mmWave-MIMO, precoding can leverage large antenna array in small area to combat path loss with smaller wavelength. However, huge cost of RF chains at high frequencies, as well as large number of antennas, makes full-digital precoding in mmWave-MIMO infeasible.

To address RF cost issue, hybrid precoding, which splits the full-digital precoder into a radio frequency (RF) precoder cascaded with a baseband precoder [2], can reduce the number of RF chains while achieving similar performance as full-digital precoder. However, hybrid precoder design in [2] requires iterative process for interference cancellation and matrix inversion for calculating BB precoder. Hence, several approaches [3]-[4] have been proposed to simplify matrix inversion by applying Schur-Banachiewicz block-wise inversion [6]. The approach in [5] uses a set of predefined orthonormal matrices to replace interference cancellation with simple correlation. However, [3]-[5] still acquire high complexity as will be explained in Section II.

In this paper, we propose a novel hybrid precoding algorithm that totally avoids the need for matrix inversion and eliminates the iterative process for interference cancellation based on a predefined codebook and orthogonal mapping process. The major contribution of this paper are:

- Propose a multilevel-DFT codebook as shown in Fig. 1, which consists of a conventional DFT matrix and a series of rotated DFT matrices.
- Propose a low-complexity hybrid precoding algorithm based on orthogonal mapping process, i.e., Orthogonal Mapping based Matching Pursuit (OMBMP) in Fig. 1, which leverages orthogonality to eliminate iterative process and matrix inversion in hybrid precoding design.

II. REVIEW OF MILLIMETER WAVE MIMO HYBRID PRECODING DESIGN

A. Notations

We use the following notations in this paper: \( \mathbf{A} \), \( \mathbf{a} \), a stand for a matrix, a column vector and a constant. \( \mathbf{A} \) is a set. \([\mathbf{A}] \) is horizontal concatenation. \( \mathbf{A}^{(i)} \) or \( \mathbf{A}(\ell) \) is the \( i \)-th column of \( \mathbf{A} \). \( \mathbf{A}^t \), \( \mathbf{A}^{-1} \), \( \|\mathbf{A}\|_F \) stand for the conjugate transpose, inverse, and Frobenius norm of \( \mathbf{A} \). \( \mathbf{A}^{[u,j]} \) or \( \mathbf{A}(\mathcal{U},\mathcal{J}) \) is the submatrix of matrix \( \mathbf{A} \) with index sets \( \mathcal{U} \) and \( \mathcal{J} \). \( |\mathcal{U}| \) is the cardinality of set \( \mathcal{U} \). \( \mathbf{I}_N \) is an \( N \times N \) identity matrix. \( \mathcal{CN}(\mathbf{a}, \mathbf{A}) \) is a complex Gaussian vector with mean \( \mathbf{a} \) and covariance matrix \( \mathbf{A} \).
B. System Model

Consider a single-user mmWave-MIMO system shown in Fig. 1. \( N_t \) data streams are sent from the base station (BS) with \( N_t \) transmit antennas and received by the user equipment (UE) with \( N_r \) receive antennas. \( N_r \times N_r \) RF chains are equipped at the BS (UE) to enable multi-stream transmission such that \( N_t \leq N_r \leq N_t \), \( \forall \delta \in \{ t, r \} \).

The hybrid precoder at BS is composed of a \( N_r \times N_r \) baseband digital precoder \((F_{bb})\) and a \( N_t \times N_r \) RF analog precoder \((F_{rf})\) which is realized by pure analog shifters with power constraints \( \|F_{bb}F_{rf}\|^2 = N_r \). Based on the sparse property of mmWave channel, we adopt a widely used 2D narrowband frequency-nonselective mmWave channel model with \( L \) scatterers \([7]\):

\[
H = \sqrt{\frac{N_t N_r}{L}} \sum_{i=1}^{L} a_i(\varphi_i^t) a_i^H(\varphi_i^r). \tag{1}
\]

\( \alpha \sim \mathcal{CN}(0, 1) \) is the complex path gain, \( \varphi_i^t \) and \( \varphi_i^r \) are the azimuth angles of departure (AoD) and arrival (AoA) of \( i \)th scatterer. \( a_i(\varphi_i^t) \) and \( a_i(\varphi_i^r) \) are the normalized array response vectors at transmitter and receiver, respectively. Assume uniform linear array (ULA) is adopted at BS and UE, the array response vector can be expressed as

\[
\begin{equation}
\begin{aligned}
a_i(\varphi_i^t) &= \frac{1}{\sqrt{N_r}}[1, e^{j2\pi \frac{\lambda d}{\lambda d 1}}, \ldots, e^{j2\pi \frac{l(l-1)\sin(\varphi_i^t)}{\lambda d}}]^T, \\
\end{aligned}
\end{equation}
\]

where \( \lambda \) and \( d \) are signal wavelength and antenna spacing, respectively. The precoded signal through the channel and received by the hybrid combiner \( (W_{bb}W_{rf}) \) at UE is

\[
y = \sqrt{\rho} W_{bb}^T W_{rf}^H H F_{bb} F_{rf} s + W_{bb}^T W_{rf}^H n. \tag{3}
\]

\( y \in \mathbb{C}^{N_r \times 1} \) is the received signal, \( \rho \) is the average path loss, \( s \in \mathbb{C}^{N_t \times 1} \) is the input signal, and \( n \sim \mathcal{CN}(0, \sigma_n^2 I_{N_r}) \) is the received noise.

C. Review of Hybrid Precoding Designs

The achievable rate with hybrid precoder and combiner \([2]\) is given by

\[
R = \log_2 |I_{N_r} + \frac{1}{N_t} \Re\{W_{bb}^T W_{rf}^H H F_{bb} F_{rf} s + W_{bb}^T W_{rf}^H n\}|, \tag{4}
\]

where \( \rho/\sigma_n^2 \) is the received SNR, and \( \Re\{\cdot\} = \sigma_n^2 W_{bb}^T W_{rf}^H H W_{bb} \) is the noise covariance matrix after combining. The objective of hybrid precoder design is to maximize the achievable rate among all possible solutions of \( (F_{bb}, F_{bb}, W_{bb}, W_{bb}) \). In this paper, we focus on the design of hybrid precoder in downlink transmission, extension to hybrid combiner is possible and straightforward \([2]\). The hybrid precoder problem is formulated as

\[
(F_{bb}^{op}, F_{rf}^{op}) = \arg \min_{F_{bb}, F_{rf}} \|F_{bb} - F_{bb}F_{rf}\|_F \quad s.t. F_{bb}(:,i) \in \{\mathbb{A}_{opt}(,k), 1 \leq k \leq L\}, \tag{5}
\]

\( i = 1, 2, \ldots, N_r \) are the largest \( N_r \) eigenvalues of \( H \). \( \mathbb{A}_{opt} \in \mathbb{C}^{N_r \times \alpha} \) is a candidate matrix consisting of \( L \) array response vectors in (2) where \( N_r \) beamforming vectors of \( F_{rf} \) are chosen from:

\[
\mathbb{A}_{opt} = [a_1(\varphi_1^t), a_1(\varphi_2^t), \ldots, a_1(\varphi_\alpha^t)]. \tag{6}
\]

Basically, \( F_{bb} \) and \( \mathbb{A}_{opt} \) can be obtained by the channel estimation method in \([7]\). In \([2]\), Eq. (5) is solved by Simultaneous Orthogonal Matching Pursuit (SOMP). The algorithm can achieve near-optimal performance with reduced number of RF chains. However, it requires iterative process for selecting \( N_r \) beamforming vectors of \( F_{rf} \) and matrix inversion in each iteration for calculating baseband precoder. To solve matrix inversion problem, Sliding Window-Index-Selection Matrix-Inversion-Bypass Simultaneous Orthogonal Matching Pursuit (SWIS-MIB-SOMP) \([4]\) is proposed to perform matrix inversion iteratively by a smaller dimension of \( m \times m \) matrix inversion with \( m \) degree of parallelism. Therefore, \( m \) can’t be too large for practical implementation. Further, Orthogonal Matching and Local Search (OM+LS) \([5]\) is proposed using a set of orthonormal matrices to replace interference cancellation with simple correlation, however, it still requires iterative process to calculate baseband precoder, and the correlation overhead increases when the number of orthonormal matrices increases. To sum up, \([2]\)-\([5]\) still have high complexity in hybrid precoder design.

III. PROPOSED LOW-COMPLEXITY HYBRID PRECODING ALGORITHM

A. Multilevel-DFT Codebook Design

In this section, we propose using rotated multilevel Discrete Fourier Transform (DFT) codebook as the candidate beamforming matrix. DFT matrix is preferred as it has the following properties: 1) The columns in DFT matrix are uncorrelated, 2) Linear combinations of columns in DFT matrix is able to synthesize the space that is spanned by array response vectors with arbitrary directions. A conventional \( N_t \times N_t \) DFT matrix is expressed as

\[
\mathbb{A}_{DFT}(,k) = \frac{1}{\sqrt{N_t}}[e^{j2\pi \frac{m(\alpha-1)}{\alpha N_t}}, \ldots, e^{j2\pi \frac{m(m-1)}{\alpha N_t}}]^T, k = 1, 2, \ldots, N_t. \tag{7}
\]

The beam pattern of \( \mathbb{A}_{DFT} \) is shown in Fig. 2(a) assuming \( N_t = 16 \). By applying DFT matrix, the BS is able to steer signals toward \( N_t \) directions. If we further rotate each DFT column by \( 2\pi(m-1)/(\alpha N_t) \) phase-shift, the DFT matrix is transformed to a new orthonormal matrix:

\[
\mathbb{A}_{DFT}(,k) = \frac{1}{\sqrt{N_t}}[e^{j2\pi \frac{(m-1)\alpha}{\alpha N_t}}, \ldots, e^{j2\pi \frac{(m-1)\alpha N_t}{\alpha N_t}}]^T, k = 1, 2, \ldots, N_t, \tag{8}
\]

where \( m = 1, 2, \ldots, \alpha \) is the phase rotation unit. The beam pattern of the rotated DFT matrix is shown in Fig. 2(b) assuming \( \alpha = 2, m = 2 \). Combining all the \( \alpha \) orthonormal matrices \( \mathbb{A}_{DFT, \alpha} \) with \( m = 1, 2, \ldots, \alpha \), we formulate a predefined multilevel-DFT codebook \( \mathbb{M} \in \mathbb{C}^{N_t \times N_t} \) as

\[
\mathbb{M}(,k) = \frac{1}{\sqrt{N_t}}[e^{j2\pi \frac{(m-1)\alpha}{\alpha N_t}}, \ldots, e^{j2\pi \frac{(m-1)\alpha N_t}{\alpha N_t}}]^T, k = 1, 2, \ldots, \alpha N_t. \tag{9}
\]
The beam pattern of $\mathbf{M}$ is shown in Fig. 2(c). Note that if we perform $\alpha$-spacing sampling of $\mathbf{M}$ with initial index $m=1,2,...,\alpha$, we will get a corresponding orthonormal matrix as $\mathbf{A}_{\text{DFT},m}$ in (8), that is,

$$Q_m = \{ \mathbf{M}(:,m), \mathbf{M}(:,m+\alpha),...,\mathbf{M}(:,m+\alpha \times (N_t-1)) \} = \mathbf{A}_{\text{DFT},m}. \quad (10)$$

With multilevel-DFT codebook $\mathbf{M}$, the BS can steer signals toward more accurate directions with more feasibility of beamforming directions, while also take RF hardware constraints into consideration.

**B. Low-Complexity Orthogonal Mapping based Matching Pursuit (OMMP)**

With multilevel-DFT codebook $\mathbf{M}$, we propose a low-complexity hybrid precoder design based on orthogonal mapping process. For practical implementation, it is feasible to assume that only a quantized information of candidate matrix $\mathbf{A}_{\text{quantized}}$ is obtained rather than infinite resolution $\mathbf{A}_{\text{inf}}$ in [2]-[5], that is,

$$\mathbf{A}_{\text{quantized}} = \{ \phi_1, \phi_2, ..., \phi_L \}, \quad (11)$$

where each phase of $\phi_l$ is quantized by limited number of bits. Let $\{ \mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_L \}$ be empty sets corresponding to $\alpha$ orthonormal matrices in (10). At first, we map each array response vector in (11) to the corresponding index $k_l$ when $\mathbf{a}_l$ has smallest Euclidean distance, this step is simplified to a quantization step according to (2), (9):

$$k_l = \text{round} \left( \frac{d \sin(\phi_l)}{\kappa} \times \alpha N_t \right).$$

Then, $\mathbf{W}_l = \{ \mathbf{V}_l[k_l] \}$, where $m = (k_l-1) \mod \alpha + 1$. The operation is shown in Fig. 3(a). After collecting the indexes with array response vectors in (11) and a corresponding sets $\{ \mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_L \}$, we can find $\mathbf{Q}_m$ with

$$n = \arg \max_{\mathbf{V}_k} \| \mathbf{V}_k \|. \quad (13)$$

Eq. (13) implies that $\mathbf{Q}_m$ may have the largest correlation with $\mathbf{A}_{\text{quantized}}$. Next, we map the array response vectors in (11) to the corresponding index of columns in $\mathbf{Q}_m$. This step is similar to (12) with $\alpha N_t$ replaced with $N_t$ to compute the corresponding set $\mathbf{W}$ as shown in Fig. 3(b).

With $\mathbf{Q}_m$ and $\mathbf{W}$, we have a new orthonormal candidate matrix $\mathbf{Q} = \{ \mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_L \}$ in hybrid precoding design in (5). The following is summarized in Algorithm 1. At first, we perform correlation between $\mathbf{Q}$ and $\mathbf{F}_{\text{opt}}$ at Step 3, where $\mathbf{Q}_m \in \mathbb{C}^{L \times N_t}$ is the correlation matrix between $\mathbf{Q}$ and $\mathbf{F}_{\text{opt}}$. The power spread on each direction is calculated at Step 4, where $\beta \in \mathbb{C}^{L \times 1}$ is the vector with $i$th entry equals to the power spread of $\mathbf{F}_{\text{opt}}$ on the $i$th beamforming direction of $\mathbf{Q}$. The $N_{\text{bb}}$ beamforming vectors of $\mathbf{F}_{\text{bb}}$ can be selected in parallel with $N_{\text{bb}}$ largest values of $\beta$ at Step 6 to 10 such that

$$\mathbf{F}_{\text{bb}} = \mathbf{Q}(:,\mathcal{V}),$$

where $\mathcal{V}$ is the set corresponding to $N_{\text{bb}}$ largest values of $\beta$ and $\mathcal{V} = \{ \mathcal{V}_1, \mathcal{V}_2, ..., \mathcal{V}_L \}$. After deriving $\mathbf{F}_{\text{bb}}$, the corresponding BS precoder $\mathbf{F}_{\text{bb}}$ is calculated by least square solution:

$$\mathbf{F}_{\text{bb}} = (\mathbf{F}_{\text{bb}}^H \mathbf{F}_{\text{bb}})^{-1} \mathbf{F}_{\text{bb}}^H \mathbf{F}_{\text{opt}}.$$

---

**Algorithm 1: Hybrid Precoder Reconstruction Using Orthogonal Mapping based Matching Pursuit (OMMP)**

**Require**: $\mathbf{F}_{\text{opt}}(\phi_1, \phi_2, ..., \phi_L)$

1. do (12), $\mathbf{W}_k = \{ k \}$, do (12) with $\alpha N_t$
2. derive the set $\mathbf{W}$
3. $\mathbf{Q} = \{ \mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_L \}$
4. $\mathbf{F}_{\text{opt}} = \mathbf{Q}(:,\mathcal{V})$
5. $\mathbf{F}_{\text{bb}} = \mathbf{F}_{\text{opt}}$
6. for $i \leq N_{\text{bb}}$
7. $k = \arg \max \| \mathbf{V}_k \|$
8. $\mathcal{V} = \{ k \}$
9. $\beta(i) = 0$
10. end for
11. $\mathbf{F}_{\text{bb}} = \mathbf{Q}(:,\mathcal{V})$
12. $\mathbf{F}_{\text{bb}} = \mathbf{F}_{\text{opt}}$
13. $\mathbf{F}_{\text{bb}} = (\mathbf{F}_{\text{bb}}^H \mathbf{F}_{\text{bb}})^{-1} \mathbf{F}_{\text{bb}}^H \mathbf{F}_{\text{opt}}$.
Since $F_{BF}$ consists of orthogonal beamforming vectors, it is a unitary matrix, therefore, the least square solution in (15) can be simplified to

$$
F_{BF} = (F_{BF}^H F_{BF})^{-1} F_{BF}^H F_{BF}^{opt} = I_{N_t} F_{BF} F_{BF}^{opt}
$$

which means that with orthonormal candidate matrix $Q$, the matrix inversion can be completely eliminated and is replaced with simple correlation. Although we adopt DFT matrix throughout this paper, any set of matrices is suitable for the proposed algorithm as long as the matrices have orthogonal property.

IV. SIMULATION RESULTS AND COMPLEXITY ANALYSIS

In this section, we present simulation results of the proposed algorithm based on multilevel-DFT codebook and orthogonal mapping process.

A. System Configuration

Assume the BS is equipped with $N_t=64$ antennas to transmit $N_s=2$ data streams and the UE is with $N_r=16$ antennas. Both BS and UE adopt ULA antennas with spacing $d=\lambda/2$. The mmWave channel is modeled by eight propagation paths ($L=8$) with AoDs and AoAs uniformly distributed between $[0,2\pi)$. This system is assumed to operate at 28 GHz carrier frequency. The spectrum efficiency is calculated based on (4) with full-digital minimum mean square error (MMSE) combiner [2] adopted at UE. All simulations are performed under 3000 channel realizations.

B. Spectrum Efficiency

Assume we have the optimal full-digital precoder $F_{opt}$, and the quantized information of candidate matrix $A_{quantized}$ with 7-bit resolution, i.e., the phases of $\mathbf{a}_k(\theta_k)$ in (11) are distributed uniformly in the set $\{2\pi k/2^7\}_{k=1,2,\ldots,2^7}$. In Fig. 4(a), we compare the performance of the proposed OMBMP, SWIS-MIB-SOMP, and OM+LS versus SNR with $\alpha=4$, $\alpha=2$, and $\alpha=2$ for pair comparison. We can observe that the proposed OMBMP can achieve 97.6% spectrum efficiency of OM+LS and achieve the same spectrum efficiency of SWIS-MIB-SOMP.

Based on the parameters in this section, Fig. 4(b) compares the performance of the proposed OMBMP with/without multilevel-DFT codebook. Without multilevel-DFT codebook means that $\alpha=1$, the algorithm will directly maps the array response vector in (11) to the corresponding vectors in ordinary DFT matrix $A_{DFT}$. We can observed that with multilevel-DFT codebook, this algorithm can select a better orthonormal matrix that steers vectors in a more precise direction, which increases 2.3% spectrum efficiency.

C. Analysis of Complexity

We compare the complexity in terms of complex multiplications in Table I. $\alpha$ is the number of orthonormal matrices in OM+LS [5]. Assume $\alpha=2$, $N_t=N$, $N_{opt}=L=\text{floor}(N/2)$, and $N_{opt}=\text{floor}(N/4)$.

Fig. 4. Achievable rate of (a) proposed OMBMP and related hybrid precoding algorithms versus SNR, and (b) proposed OMBMP with/without multilevel-DFT codebook versus SNR.

Fig. 5. Computational complexity of (a) proposed OMBMP and OM+LS, and (b) proposed OMBMP and SWIS-MIB-SOMP.
In this paper, we have presented a low-complexity hybrid precoding algorithm for mmWave-MIMO communication systems. The proposed algorithm is performed based on a predefined multilevel-DFT codebook and orthogonal mapping process. Simulation results demonstrate that the proposed algorithm can achieve almost the same spectrum efficiency compared with state-of-the-art low-complexity hybrid precoding algorithms with greatly reduced complexity.

ACKNOWLEDGMENT

This work was financially supported by the Ministry of Science and Technology of Taiwan under Grants MOST 103-2622-E-002-034 & 104-2622-8-002-002, and sponsored by MediaTek Inc., Hsin-chu, Taiwan, and NOVATEK Fellowship.

REFERENCES


