An Enhanced Nonlinear Hybrid Precoding Algorithm for Millimeter Wave Massive MIMO Systems

Fulai Liu¹, Yanshuo Zhang¹* and Ruiyan Du¹

¹Engineer Optimization & Smart Antenna Institute, North eastern University at Qinhuangdao, Hebei, 066000, China
*Corresponding author’s e-mail: shuoshuo012345@126.com

Abstract. Hybrid precoding has recently received considerable attentions in millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) systems. Especially, a nonlinear hybrid Tomlinson-Harashima precoding (THP) algorithm has recently received considerable attentions since it can effectively improve the performance of the hybrid precoding. This paper presents an enhanced nonlinear hybrid Tomlinson-Harashima precoding algorithm based on the column-by-column refinement (named as E-THP-CR algorithm). Compared with the previous works, the proposed algorithm offers a number of advantages. The advantage of the proposed algorithm lies in achieving the column-by-column refinement of the analog precoding matrix. Afterwards, the digital precoding matrix is decomposed into two separate sections which respectively belong to the encoding part and effective channel part. This design can effectively remove the inter-user interference as expected in the THP module. Simulation results verify that the proposed algorithm provides higher spectral efficiency, compared with the other relevant algorithms.

1. Introduction

Millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) technology is emerging as a promising technology in the Fifth Generation Communication Technology (5G) for the enormous idle spectrum resources within the mmWave bands [1]. In traditional MIMO system, the precoding is achieved fully digitally at the baseband, which demands that each antenna has a dedicated radio frequency (RF) chain. However, the prohibitive cost and power consumption of radio frequency (RF) chains make fully digital precoding infeasible at mmWave bands. Consequently, a hybrid precoding architecture which can offer comparable performance with few RF chains compared with fully digital precoding has been proposed recently [2]. The hybrid precoding can be divided into two categories: linear hybrid precoding and nonlinear hybrid precoding. However, the linear hybrid precoding suffers from the significant performance loss. Fortunately, this phenomenon could be effectively relieved by the nonlinear hybrid precoding [3].

Dirty paper coding (DPC) is considered as the optimal nonlinear precoding algorithm for multi-user MIMO downlink channel [4]. Nevertheless, it is unfeasible in practice due to the high computational cost. Tomlinson-Harashima precoding (THP) is a low complexity and suboptimal implementation of the DPC. Therefore, the THP algorithm has received significant attentions in recent years. A hybrid Tomlinson-Harashima precoding algorithm based on the vector perturbation (TH-VP) is proposed in [5], which combines THP and VP to nullify both inter-user and inter-stream interference. A low complexity nonlinear hybrid Tomlinson-Harashima precoding algorithm based on the block diagonal geometric mean decomposition (BD-GMD-THP) is proposed in [6], in which the
hybridBD-GMD-THP matrix is obtained by the orthogonal matching pursuit (OMP) algorithm, and then the digital precoding matrix is reconstructed by the decomposition of analog precoding matrix. To further achieve a better performance of the hybrid precoding, in this paper, an enhanced nonlinear hybrid Tomlinson-Harashima precoding algorithm based on the column-by-column refinement (named as E-THP-CR algorithm) is proposed. In the proposed algorithm, the initialized hybrid precoding matrix is obtained via the OMP algorithm. Then, all column vectors of the analog precoding matrix are refined column by column. Afterwards, the refined analog precoding matrix is used as the new input matrix for the next iteration to update the hybrid precoding matrix via the OMP algorithm again. Finally, the digital precoding matrix is decomposed into two separate sections which respectively belong to the encoding part and effective channel part, in order to effectively remove the inter-user interference as expected in the THP module.

2. Datamodel
In this section, a MU-MIMO downlink model for mmWave communication system is adopted as shown in figure 1, in which a base station (BS) is equipped with $N_t$ transmitting antennas and $N_{rf}$ RF chains. The BS is assumed to serve $K$ users equipped with $N_r$ receiving antennas and one RF chain. $N_t$ stands for the number of data streams.

![Figure 1. Multi-user mmWave MIMO system with hybrid precoding and combining](image)

In order to ensure the quality of communication, it is necessary to satisfy $N_r \leq N_t$. The transmitted signal $x$ at the BS can be written as

$$x = F_{RF} F_{BB} s$$

where $F_{RF} \in \mathbb{C}^{N_t \times N_r}$ represents the analog precoding matrix, and $F_{BB} \in \mathbb{C}^{N_r \times N_t}$ stands for the digital precoding matrix. $s \in \mathbb{C}^{N_r \times 1}$ stands for the transmitted signal. The analog precoding matrix $F_{RF}$ and the digital precoding matrix $F_{BB}$ are designed to meet the total transmitting power constraint $\|x\|_2^2 \leq P_t$, where $P_t$ denotes the constrained total power at the BS.

The received signal $y_k$ for the $k^{th}$ user after combining can be represented as

$$y_k = w_{BB} w_{RF}^H H_k F_{RF} F_{BB} s + w_{BB} w_{RF}^H n_k$$

where $H_k \in \mathbb{C}^{N_r \times N_t}$ denotes the channel matrix between the BS and the $k^{th}$ user. $w_{RF} \in \mathbb{C}^{N_r \times 1}$ and $w_{BB}$ stand for the analog RF combining vector and scaling factor at the receiver, respectively. The noise signal $n_k \in \mathbb{C}^{N_r \times 1}$ follows the Gaussian distribution. $\sigma_n^2$ represents the noise power.

Due to $F_{RF}$ and $w_{RF}$ are processed only by phase shifters, all the elements in $F_{RF}$ and $w_{RF}$ are limited by the unit modulus constraint, namely, $\|F_{RF}\|_\infty = 1, \|w_{RF}\|_\infty = 1$.

The Rician channel model is considered in this paper, which consists of LOS and NLOS components. Thus, the channel matrix of the $k^{th}$ user $H_k$ can be described as follows [7].
where the LOS component and the NLOS component of \( \mathbf{H} \) is described as \( \mathbf{H}_{\text{LOS},k} \) and \( \mathbf{H}_{\text{NLOS},k} \), respectively. \( v_k \) represents the Rician factor.

The NLOS component \( \mathbf{H}_{\text{NLOS},k} \) is assumed to be a sum of \( N_{\alpha,k} \) clusters, each cluster consists of \( N_{\alpha,k} \) paths. Consequently, the LOS component \( \mathbf{H}_{\text{LOS},k} \) and the NLOS component \( \mathbf{H}_{\text{NLOS},k} \) can be written as[7] \( H_{\text{LOS},k} = \alpha_k a_k (\phi_k', \theta_k') a_k (\phi_k', \theta_k')^H \) \( (4) \)

\[
H_{\text{NLOS},k} = \frac{1}{\sqrt{N_{\alpha,k} N_{\text{NLOS},k}}} \sum_{i=1}^{N_{\alpha,k}} \sum_{l=1}^{N_{\text{NLOS},k}} \alpha_{\alpha,i,l} a_{\alpha,i,l} (\phi_{\alpha,i,l}', \theta_{\alpha,i,l}') a_k (\phi_k', \theta_k')^H \]
\( (5) \)

where \( \alpha_\alpha \) denotes the complex gain in the LOS path, and \( \alpha_{\alpha,i,l} \) denotes the complex gain for the \((i,l)\)th path of the NLOS component. \( \phi_\alpha'(\theta_\alpha') \) and \( \phi_\alpha'(\theta_\alpha') \) are described as the azimuth (elevation) angle of arrival (AOA) and the azimuth (elevation) angle of departure (AOD) of the LOS path. \( \phi_{\alpha,i,l}(\theta_{\alpha,i,l}) \) and \( \phi_{\alpha,i,l}(\theta_{\alpha,i,l}) \) stands for AOA and AOD of the \((i,l)\)th path of the NLOS component respectively. \( a_\alpha (\phi_\alpha', \theta_\alpha') \) and \( a_k (\phi_k', \theta_k') \) denote the array response vectors of BS and user, respectively. For an \( M \times N \) uniform planar array (UPA) in the \( yz \)-plane with a half-wavelength inter-element space, the array response vector can be expressed as[8]

\[
a(\phi, \theta) = [1, \ldots, e^{j(M-1)\sin\phi \sin\theta + (N-1)\cos\theta}, \ldots, e^{j(M-1)\sin\phi \sin\theta + (N-1)\cos\theta}]^T
\]
\( (6) \)

where \((0 \leq m \leq M-1)\) and \((0 \leq n \leq N-1)\) are the \( y \) and \( z \) indices of array element, respectively.

### 3. Algorithm formulation

In this section, a nonlinear hybrid E-THP-CR algorithm is proposed. In proposed algorithm, the precoding and combining problem are reformed as a sparse signal reconstitution. Then, the precoding problem and combining problem can be expressed as [6]

\[
\arg \min_{F_{\text{opt}}, F_{\text{opt}}} \| F_{\text{opt}} - F_{\text{opt}}F_{\text{opt}} \|_{F}^2
\]
\( s.t. \| F_{RF(m,n)} \|_{F} = 1, \forall m, n \)
\( (7) \)

where \( F_{\text{opt}} \) is the fully digital precoding matrix, and \( \| F_{RF} \|_{F} = N_z \) denotes the normalized transmitting power constraint.

\[
\arg \min_{\text{w}_{\text{opt}}, \text{w}_{\text{opt}}} \| W_{\text{opt}} - W_{\text{opt}}W_{\text{opt}} \|_{F}^2
\]
\( s.t. \| W_{RF(k,i)} \|_{F} = 1, \forall k,i \)
\( (8) \)

where \( W_{\text{opt}} \) is the combing matrix for the \( k^{th} \) user at the receiver, \( W_{RF} \) and \( W_{RF} \) stand for the scaling factor and the analogRF combining vector for the \( k^{th} \) user at the receiver, respectively.

The fully digital precoding matrix \( F_{\text{RF}} \) and the combing matrix \( W_{\text{RF}} \) can be obtained according to [9]. In proposed algorithm, the digital precoding matrix \( F_{\text{RF}} \) is decomposed into a product of the matrices \( F_{\text{RF}} \) and \( F_{\text{RF}} \), in order to remove the inter-user interference. Then the diagram of hybrid precoding is given as in figure 2, in which the matrix \( F_{\text{RF}} \) is exploited to implement the digital THP algorithm and the matrix \( F_{\text{RF}} \) is used as the part of the effective channel.

The OMP algorithm can be used as an approximate solution for (7) and (8). It is worth noting that the OMP algorithm depends on the predefined codebook. When the predefined code books notac-curate enough, the performance of the hybrid precoding suffers from the significant loss.
The initialized hybrid precoding matrix can be obtained via the OMP algorithm, and the analog precoding matrix is refined in order to reduce the residual \( \beta \) between the fully digital precoding matrix and the hybrid precoding matrix. The residual \( \beta \) can be expressed as [10]

\[
\beta = \|F_{\text{opt}} - F_{RF}F_{BB}\|_F - \sum_{i=1}^{N_F} f^i_{RF}f^i_{BB}\|_F
\]

where \( f^i_{RF}(1 \leq i \leq N_F) \) and \( f^i_{BB}(1 \leq i \leq N_F) \) denote the \( i^{th} \) column of the analog precoding matrix and the \( i^{th} \) row of the digital precoding matrix, respectively.

The objective function of the refining process can be described as [10]

\[
\arg \min_{F_{RF(n,m)}} \| E_u - f^i_{RF}f^i_{BB}\|_F = \| E_u - \sigma_i u_i \|_F
\]

where \( E_u = F_{\text{opt}} - \sum_{i=1}^{N_F} f^i_{RF}f^i_{BB} \), \( \sigma_i \) is the largest singular value of \( E_u \), \( u_i \) and \( v_i \) are the first left and the first right singular vector of \( E_u \).

After all columns of the initial hybrid precoding matrix have been refined, the modified analog precoding matrix is used to be the new input matrix for the next iteration to update the hybrid precoding matrix via the OMP algorithm again. The distortion is defined as the residual \( \beta \). When the difference between the adjacent distortion values is less than the given threshold, the proposed algorithm stops the iteration, and outputs the updated hybrid precoding matrix. Considering the unit modulus constraint, the constrained analog precoding matrix can be given by [11]

\[
F_{RF(n,m)} = \frac{\hat{F}_{RF(n,m)}}{|\hat{F}_{RF(n,m)}|}
\]

where \( \hat{F}_{RF} \) stands for unconstrained updated analog precoding matrix.

In order to simplify the baseband precoder design problem, perform QR decomposition on the analog precoding matrix \( F_{RF} \) [12]

\[
F_{RF} = Q_{RF}R_{RF}
\]

where \( Q_{RF} \) is a semi-unitary matrix, \( R_{RF} \) is a upper-triangular matrix, \( R_{RF}^{-1} = F_{RF} \).

After the constrained modified analog precoding matrix can be obtained, the effective channel matrix for the \( k^{th} \) user between the transmitter and the receiver \( H_{\text{eff}} \) can be described as

\[
H_{\text{eff}} = w^H_{RF}H_{RF}F_{RF} = w^H_{RF}H_{RF}Q_{RF} = L_{RF}Q_{RF}
\]

where \( H_{RF} = [H_{RF1}, \ldots, H_{RFk}]^T \) can be defined as the whole effective channel matrix.

The matrix \( F_{RF} \) is derived by the BD-GMD-ss decomposition of the whole effective channel matrix [13]

\[
P_{RF}H_{\text{eff}}Q_{RF} = L_{RF}
\]

where \( P_{RF} = \text{Blkdiag}(P_{RF1}, \ldots, P_{RFk}) \); \( P_{RF} \) and \( Q_{RF} \) are both semi-unitary matrices; \( L_{RF} \) is a lower-triangular matrix whose diagonal block contains identical diagonal element; \( A_{RF} = \text{Diag}(\text{diag}(L_{RF})) \).
Then, the matrix $F_D$ can be described as [14]

$$F_D = \gamma_D Q_D A_D^\dagger$$  \hfill (15)

where $\gamma_D$ is the power scaling factor.

The digital precoding matrix $F_{bb}$ can be written as

$$F_{bb} = F_r F_D$$ \hfill (16)

where the digital precoding matrix $F_{bb}$ is reconstructed via the matrices $F_r$ and $F_D$.

4. Simulation results

In this section, some simulation results on spectral efficiency will be presented to evaluate the performance of the proposed hybrid precoding for multi-user mm Wave MIMO systems. All terminals are equipped with UPA. The Rician fading channel is considered in this section, the value of the Rician factor $\nu$ is uniformly distributed $[1,10]$. The AOA and AOD of the LOS component $H_{LOS,k}$ are uniformly distributed in $[0,2\pi)$. The number of clusters in the NLOS component $H_{NLOS,k}$ is assumed as $N_{i,k} = 5$. The number of the paths in each cluster is set up as $N_{i,k} = 10$. The average AOA and AOD of the $i$th cluster are uniformly modelled as $[0,2\pi)$. The AOA and AOD of the $i$th ray within the $i$th cluster $\phi_{i,k,l}$, $\theta_{i,k,l}$ and $\alpha_{i,k,l}$ follow the Laplace distribution with the scale parameter being $10^\circ$. The complex gain $\alpha_{i,k,l}$ of the $i$th ray within $i$th cluster is set up as a $CN(0,1)$ random variable. All simulation results can be averaged over 1000 channel realizations.

In this section, the performance of the proposed algorithm is compared with the optimal result given by the cooperative users, the linear hybrid BD precoding algorithm in [4], and the nonlinear hybrid BD-GMD-THP algorithm in [6]. Figure 3 shows the spectral efficiency of the proposed method in terms of different signal-to-noise ratio (SNR). From this figure, it can be observed that the spectral efficiency of the proposed method outperforms the other hybrid precoding algorithms. This phenomenon depends on the refinement of the analog precoding matrix, which can effectively avoid the performance loss of the hybrid precoding. Figure 4 and Figure 5 present the spectral efficiency comparison with the different algorithms as functions of $N_t$ and $N_r$, respectively. As can be observed in the two figures, the spectral efficiency of all algorithms is improved with the increasing of the number of antennas. It is can be seen that the efficiency of the proposed algorithm is closer to the optimal spectral efficiency than the other algorithms. With the further observation, the performance of the proposed algorithm remains almost stable. Meanwhile, it implies that the proposed algorithm is still adapted to the massive antenna arrays.
5. Conclusions
This paper presents an enhanced nonlinear hybrid THP algorithm based on the column-by-column refinement. In the proposed algorithm, the analog precoding matrix is firstly refined column by column. Afterwards, the hybrid precoding matrix is updated via the OMP algorithm again. Finally, the digital precoding matrix is split into two separate parts, in order to improve the spectral efficiency. The results in simulation demonstrate that the proposed algorithm can achieve the better performance of the hybrid precoding, compared with the other relevant algorithms.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (Grant No.61971117) and by the Fundamental Research Funds for the Central Universities (Grant No.N172302002).

References
[1] Li, S.D., Liu, Y.J., Lin, L.K., Sun, X.C., Yang, S., Sun, D.M. (2018) Millimeter-Wave channel simulation and statistical channel model in the cross-corridor environment at 28 GHz for 5G wireless system. In: International Conference on Microwave and Millimeter Wave Technology. Chengdu, pp. 1-3.
[2] Liu, X.R., Li, X.M., Cao, S., Deng, Q.Y., Ran, R., Nguyen, K., Yingri, P. (2019) Hybrid precoding for massive mmWave MIMO systems. J. IEEE Access., 7: 33577-33586.
[3] Peel, C.B., Hochwald, B.M., Swindlehurst, A.L. (2005) A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization. J. IEEE Transactions on Communications., 53: 195-202.
[4] Wang, Z.H., Li, M., Liu, Q., Swindlehurst, A.L. (2018) Hybrid precoder and combiner design with low resolution phase shifters in mmWave MIMO systems. J. IEEE Journal of Selected Topics in Signal Processing., 12: 256-269.
[5] Chen, R., Moretti, M., Wang, X.D. (2017) A hybrid TH-VP precoding for multiuser MIMO. J. IEEE Transactions on Vehicular Technology., 66: 11399-11403.
[6] Chang, T.Y., Chen, C.E. (2018) A hybrid Tomlinson-Harashima transceiver design for multiuser mmWave MIMO systems. J. IEEE Wireless Communications Letters., 7: 118-121.
[7] Zhao, L., Ng, D.W.K., Yuan, J.H. (2017) Multi-user precoding and channel estimation for hybrid millimeter wave systems. J. IEEE Journal on Selected Areas in Communications., 35: 1576-1590.
[8] Balanis, C. (1997) Antenna Theory. Wiley-Interscience, New York.
[9] Lin, S.W., Ho, W.W.L., Liang, Y.C. (2008) Block diagonal geometric mean decomposition (BD-GMD) for MIMO broadcast channels. J. IEEE Transactions on Wireless Communications., 7: 2778-2789.
[10] Du, R.Y., Liu, F.L., Wang, X.W., Zhou, Q.P., Bai, X.Y. (2018) P-OMP-IR algorithm for hybrid precoding in millimeter wave MIMO systems. J. Progress In Electromagnetics Research M., 68: 163-171.
[11] Alkhateeb, A., Heath, R.W. (2016) Frequency selective hybrid precoding for limited feedback millimeter wave systems. J. IEEE Transactions on Communications., 64: 1801-1818.
[12] Zhang, J., Haardt, M., Soloveychik, I., Wiesel, A. (2016) A channel matching based hybrid analog-digital strategy for massive multi-user MIMOdowntlink systems.In: IEEE Sensor Array Multichannel Signal Process Workshop (SAM).Rio de Janerio. pp. 1-5.
[13] Ho, W.W.L., Liang, Y. (2008) User Ordering and Subchannel Selection for Power Minimization in MIMO Broadcast Channels using BD-GMD. In: IEEE 68th Vehicular Technology Conference. Calgary. pp. 1-5.
[14] Ni, W., Dong, X. (2016) Hybrid block diagonalization for massive multiuserMIMO systems. J. IEEE Transactions on Communications., 64: 201–211.