Received: 12 March 2019 | Revised: 28 August 2019 | Accepted: 16 September 2019

DOI: 10.4218/etrij.2019-0123

ORIGINAL ARTICLE

ETRI Journal WILEY

Hybrid combiner design for downlink massive MIMO systems

Bangwon Seo

Division of Electrical, Electronics and Control Engineering, The Institute of IT Convergence Technology, Kongju National University, Cheonan, Rep. of Korea

Correspondence
Bangwon Seo, Division of Electrical, Electronics and Control Engineering, The Institute of IT Convergence Technology, Kongju National University, Cheonan, Rep. of Korea.
Email: seobw@kongju.ac.kr

Funding information
Basic Science Research Program through the National Research Foundation of Korea (NRF); Ministry of Education, Grant/Award Number: 2016R1D1A3B03935210

We consider a hybrid combiner design for downlink massive multiple-input multiple-output systems when there is residual inter-user interference and each user is equipped with a limited number of radio frequency (RF) chains (less than the number of receive antennas). We propose a hybrid combiner that minimizes the mean-squared error (MSE) between the information symbols and the ones estimated with a constant amplitude constraint on the RF combiner. In the proposed scheme, an iterative alternating optimization method is utilized. At each iteration, one of the analog RF and digital baseband combining matrices is updated to minimize the MSE by fixing the other matrix without considering the constant amplitude constraint. Then, the other matrix is updated by changing the roles of the two matrices. Each element in the RF combining matrix is obtained from the phase component of the solution matrix of the optimization problem for the RF combining matrix. Simulation results show that the proposed scheme performs better than conventional matrix-decomposition schemes.

KEYWORDS
hybrid combiner, limited RF chains, massive MIMO, mean-squared error (MSE)

1 | INTRODUCTION

Massive multiple-input multiple-output (MIMO), millimeter wave (*mm*Wave), and non-orthogonal multiple access (NOMA) technologies are considered essential for achieving high capacity and spectral efficiency in fifth generation (5G) wireless communication systems [1,2]. The main feature of *mm*Wave communications is the use of *mm*Wave frequency bands, which are less occupied than the frequency bands utilized by current cellular communications [3,4]. The main principle of NOMA is to allow multiple user equipment (UE) to share the same time-frequency resources without spatial separation. In NOMA, inter-user interference in the same time-frequency resources is eliminated using superposition techniques at the transmitter and successive interference cancellation techniques at the receiver [5–8]. Moreover, in massive MIMO, a large antenna array containing dozens or even hundreds of antenna elements is deployed at the base station (BS) to increase the spatial multiplexing/diversity gain significantly and to achieve very high capacity using multiuser MIMO (MU-MIMO) precoding [9–19]. Among the key technologies, in this paper, we focus on massive MIMO systems.

In traditional massive MIMO systems, the number of radio frequency (RF) chains is assumed to be the same as the number of antennas, and the processing performed for signal transmission and reception at the baseband is fully digital. This enables control of both the amplitude and phase of the incoming signal. However, a very large number of RF chains will result in high-implementation cost and power consumption, which makes it infeasible to implement massive MIMO in mobile wireless communication systems [20–22].

To address this problem, the authors in [23–29] considered hybrid processing for massive MIMO systems with the number of RF chains less than the number of antennas. Hybrid processing involves a cascaded structure of analog RF processing and digital baseband processing. To handle the mismatch

between the numbers of RF chains and antennas, high-dimensional analog RF processing was employed using cost-effective phase shifters to control only the phase of the incoming signal. Digital baseband processing capable of adjusting both the amplitude and phase of the incoming signal was performed in a very low dimension. Reducing the number of RF chains in the hybrid processing structure leads to lower implementation cost and less power consumption, compared with those of a conventional full-RF-chain configuration. However, as the combining matrix for analog RF processing needs to be designed so that all its elements have the same amplitude [23–29], the design of hybrid combining matrices is a nonconvex optimization problem and is challenging to solve.

To design the hybrid precoder and combiner in massive MU-MIMO systems, the authors in [28] designed an RF combining matrix to harvest the array gain by selecting some columns of the discrete Fourier transform matrix. With the designed RF combining matrix, a baseband combining matrix of low dimension was designed based on the singular value decomposition (SVD) of the desired channel matrix while assuming that inter-user interference was perfectly cancelled at the receiver. However, this scheme showed poor performance when there was residual inter-user interference at the receiver.

The authors in [29] and [30] presented another method for designing the analog RF and digital baseband combining matrices based on the Frobenius distance from the unconstrained minimum mean-squared error (MMSE) combining matrix. In these studies, without considering a nonconvex constraint—that is, a constant amplitude constraint—on the RF combining matrix, an unconstrained MMSE combining matrix was obtained. This was done by solving a convex optimization problem that minimized the mean-squared error (MSE) between the information symbols and their estimates. Then, two matrices each for RF and baseband combining were designed to minimize the Frobenius distance between the unconstrained MMSE combining matrix and the product of the two matrices while considering the nonconvex constraint. This nonconvex optimization problem was solved using an iterative alternating optimization method. However, this approach does not guarantee that the obtained hybrid combining matrices minimize the MSE between the information symbols and their estimates, although they might be the closest to the unconstrained MMSE combining matrix in terms of the Frobenius distance. This is especially true when there is residual inter-user interference at the receiver because of imperfect channel state information (CSI) at the transmitter (CSIT).

In this paper, we present a hybrid combiner design scheme for downlink massive MIMO systems with a reduced number of RF chains when there is residual inter-user interference caused by imperfect CSIT. In the proposed scheme, RF and baseband combining matrices are designed to minimize the MSE between the information symbols and their estimates using an iterative alternating optimization method. In particular, one of the RF and baseband combining matrices is updated to minimize the MSE without considering the constant amplitude constraint by fixing the other matrix. Then, the other matrix is updated by changing the roles of the two matrices. This alternating optimization is performed iteratively until a given condition is satisfied. In updating the RF combining matrix, all the elements of the RF combining matrix are set to have a constant amplitude, and their phases are obtained from those of the MSE solution matrix to consider the constant amplitude constraint on the RF combining matrix. Simulation results show that the proposed scheme achieves better performance than the conventional schemes in the high signal-to-noise ratio (SNR) region in terms of MSE, achievable rate, and bit-error rate (BER). Its performance is close to that of a fully digital unconstrained MMSE receiver.

The main contributions of this work can be summarized as follows:

1. We propose a hybrid combiner design using an MSE-based cost function between the information symbols and their estimates when the number of RF chains is less than the number of antennas. The conventional schemes in [29] and [30] use a Frobenius-distance-based cost function between the unconstrained MMSE combining matrix and its estimate. Consequently, the proposed scheme has lower MSE and BER performance than those of the conventional schemes.
2. An alternating optimization method is presented to solve the formulated nonconvex optimization problem with the same computation complexity as the conventional schemes in [29] and [30].

The rest of the paper is organized as follows: Section 2 describes the system model and the proposed two-step hybrid combiner is presented in Section 3. Simulation results are presented in Section 4 and the paper is concluded in Section 5.

Notations: The field of complex numbers is represented by C. The (m, n)-th element of matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{m,n}$. If the polar representation of $[\mathbf{A}]_{m,n}$ for all (m,n) is denoted by $a_{m,n}e^{j\phi_{m,n}}$, the expression $\angle\mathbf{A}$ represents a matrix whose (m, n)-th element is $\angle[\mathbf{A}]_{m,n} = e^{j\phi_{m,n}}$. An identity matrix of size $N \times N$ is denoted by $\mathbf{I}_N$. The notation $|x|$ for any scalar x indicates the amplitude of x, $\delta(n)$ is the Kronecker delta function, and $E[\cdot]$ denotes an expectation operation.

2 | SYSTEM MODEL

For downlink massive MIMO systems, to reduce implementation cost and power consumption at the UE, we consider a UE receiver structure design with a limited number of RF chains (less than the number of receive antennas).

Figure 1 shows the hybrid UE receiver structure with a limited number of RF chains. As shown in the figure, the receiver processing involves two-stage combiners: an

analog RF combiner and a digital baseband combiner. The UE is equipped with N_r receive antennas and M_r RF chains and we assume $M_r < N_r$. The RF combiner and baseband combiner are denoted by the matrices $\mathbf{W}_R \in C^{N_r \times M_r}$ and $\mathbf{W}_B \in C^{M_r \times N_s}$, respectively, where N_s denotes the number of data streams simultaneously transmitted for the UE.

We assume that a BS with N_t transmit antennas serves K UEs using the same time-frequency resources, that each UE is equipped with multiple receive antennas, and that UE 1 is a desired UE. The received signal vector $\mathbf{y}_1 \in C^{N_r \times 1}$ at the desired UE is given by

$$\mathbf{y}_1 = \mathbf{H}_1\mathbf{F}_1\mathbf{s}_1 + \sum_{k=2}^{K} \mathbf{H}_1\mathbf{F}_k\mathbf{s}_k + \mathbf{z}_1, \tag{1}$$

where $\mathbf{s}_k \in C^{N_s \times 1}$ is the symbol vector of UE k with $E[\mathbf{s}_m\mathbf{s}_n^H] = \sigma_s^2\mathbf{I}_{Ns}\delta(m-n)$, and σ_s^2 is the signal power allocated to each symbol of a UE. The matrix $\mathbf{F}_k \in C^{N_t \times N_s}$ is a transmit precoding matrix for UE k, and $\mathbf{H}_1 \in C^{N_r \times N_t}$ is a channel matrix between the BS and UE 1. Moreover, the vector $\mathbf{z}_1 \in C^{N_r \times 1}$ is an additive white complex Gaussian noise (AWGN) vector whose elements are from a distribution $CN(0, \sigma_z^2)$.

In time-division duplex systems, the CSIT at the BS is obtained using the channel reciprocity between the downlink and uplink. However, CSIT estimation error is inevitable because of integrated effects, such as estimation delay, background noise, pilot interference, and synchronization errors in the RF chains between the downlink and uplink. Therefore, the estimated CSIT for UE 1 at the BS can be written as $\hat{\mathbf{H}}_1 = \mathbf{H}_1 - \mathbf{H}_{1,e}$, where $\mathbf{H}_1$ is a true channel matrix and $\mathbf{H}_{1,e}$ is a CSIT estimation error matrix. Each element of $\mathbf{H}_{1,e}$ is modeled as a random variable with the distribution $CN(0, \sigma_e^2)$ [31]. We assume that $\mathbf{F}_k$ for all k except $k = 1$ is designed to be orthogonal to the estimated channel $\hat{\mathbf{H}}_1$ as in [29] or [30] (that is, $\hat{\mathbf{H}}_1\mathbf{F}_k = \mathbf{0}_{Nr \times Ns}$) for all $k \neq 1$. Please refer to Appendix for the detailed explanation on $\mathbf{F}_k$ design. Therefore, we know that $\mathbf{H}_1\mathbf{F}_k = \mathbf{H}_{1,e}\mathbf{F}_k$, $k = 2, 3, ..., K$.

Now, the received signal vector can be rewritten as

$$\mathbf{y}_1 = \mathbf{H}_1\mathbf{F}_1\mathbf{s}_1 + \sum_{k=2}^{K} \mathbf{H}_{1,e}\mathbf{F}_k\mathbf{s}_k + \mathbf{z}_1, \tag{2}$$

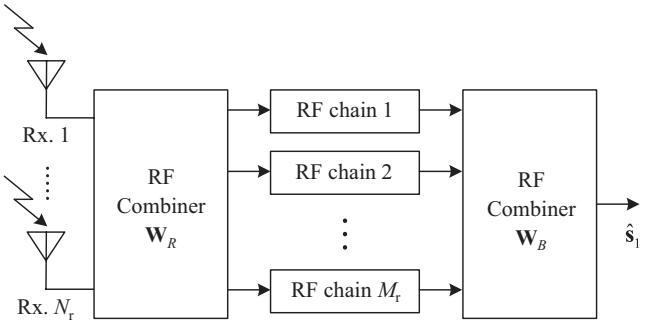

FIGURE 1 Hybrid receiver structure of the desired UE with reduced number of RF chains

where the second term is residual inter-user interference.

Let us define $\mathbf{G}_k = \mathbf{H}_1\mathbf{F}_k$. Then, we have $\mathbf{G}_k = \mathbf{H}_{1,e}\mathbf{F}_k$, $k = 2, 3, ..., K$, and the received signal can be rewritten as

$$\mathbf{y}_1 = \mathbf{G}_1\mathbf{s}_1 + \sum_{k=2}^{K} \mathbf{G}_k\mathbf{s}_k + \mathbf{z}_1, \tag{3}$$

where $\mathbf{G}_1$ is an effective channel matrix of UE 1.

We assume that UE 1 knows the effective channel matrix $\mathbf{G}_1$ perfectly. The received signal vector $\mathbf{y}_1$ is processed by the RF combining matrix $\mathbf{W}_R$ and then by the baseband combining matrix $\mathbf{W}_B$. As $\mathbf{W}_R$ is implemented by analog phase shifters, all the elements of $\mathbf{W}_R$ are constrained to have a constant amplitude such that $|[\mathbf{W}_R]_{m,n}| = 1/\sqrt{N_r}$. Thus, the output signal after two-stage combining is given by

$$\begin{aligned}\hat{\mathbf{s}}_1 &= \mathbf{W}_B^H\mathbf{W}_R^H\mathbf{y}_1 \\ &= \mathbf{W}_B^H\mathbf{W}_R^H\mathbf{G}_1\mathbf{s}_1 + \sum_{k=2}^{K} \mathbf{W}_B^H\mathbf{W}_R^H\mathbf{G}_k\mathbf{s}_k + \mathbf{W}_B^H\mathbf{W}_R^H\mathbf{z}_1.\end{aligned} \tag{4}$$

In the previous literature [29,30], the pair of hybrid combining matrices $(\mathbf{W}_R, \mathbf{W}_B)$ is designed to minimize the Frobenius distance given by

$$\min_{\mathbf{W}_R, \mathbf{W}_B} \| \mathbf{W}_o - \mathbf{W}_R\mathbf{W}_B \|_F^2 \quad \text{s.t.} \mathbf{W}_R \in \mathbf{Q}_R, \tag{5}$$

where $\mathbf{Q}_R$ is the set of matrices with all constant amplitude entries, which is $1/\sqrt{N_r}$, and $\mathbf{W}_o$ is the unconstrained MMSE combining matrix obtained by solving the following unconstrained convex optimization problem:

$$\min_{\mathbf{W}_o} E\left[\| \mathbf{s}_1 - \mathbf{W}_o^H\mathbf{y}_1 \|^2 \right]. \tag{6}$$

The nonconvex optimization problem in (5) is solved using an iterative alternating optimization method [29,30]. However, as the cost function for designing $(\mathbf{W}_R, \mathbf{W}_B)$ in the previous literature is based on the Frobenius distance $\| \mathbf{W}_o - \mathbf{W}_R\mathbf{W}_B \|_F^2$, it cannot guarantee optimality in terms of the MSE between the information symbols and their estimates, which is vital in communication systems.

3 | PROPOSED TWO-STEP HYBRID COMBINER

In this section, we propose a hybrid combiner design method based on the MSE cost function. The pair of hybrid combining matrices $(\mathbf{W}_R, \mathbf{W}_B)$ can be designed by formulating a

nonconvex optimization problem to minimize the MSE between the information symbols and their estimates given by

$$\min_{\mathbf{W}_R,\mathbf{W}_B} E\left[\left\|\mathbf{s}_1-\mathbf{W}_B^H\mathbf{W}_R^H\mathbf{y}_1\right\|^2\right] \quad \text{s.t.} \quad \mathbf{W}_R\in\mathbf{Q}_R. \tag{7}$$

This original optimization problem is not tractable because of the presence of a nonconvex constraint $\mathbf{W}_R\in\mathrm{W}_R$. To address this optimization problem, we consider the iterative alternating optimization of two unconstrained optimization problems.

We denote the hybrid combiner at the j-th iteration ($\mathbf{W}_R^{(j)}$, $\mathbf{W}_B^{(j)}$). To update the analog RF combining matrix to $\mathbf{W}_R^{(j+1)}$ for a given digital baseband combining matrix $\mathbf{W}_B^{(j)}$, we first solve the following unconstrained convex MSE minimization problem without considering the nonconvex constraint

$$\overline{\mathbf{W}}_R^{\text{opt}}=\arg\min_{\overline{\mathbf{W}}_R} E\left[\left\|\mathbf{s}_1-\mathbf{W}_B^{(j)H}\overline{\mathbf{W}}_R^H\mathbf{y}_1\right\|^2\right]. \tag{8}$$

The unconstrained optimal solution for $\overline{\mathbf{W}}_R$ is given by [32]

$$\overline{\mathbf{W}}_R^{\text{opt}}=\sigma_s^2\mathbf{R}_y^{-1}\mathbf{G}_1(\mathbf{W}_B^{(j)H}\mathbf{W}_B^{(j)})^{-1}\mathbf{W}_B^{(j)H}, \tag{9}$$

where $\mathbf{R}_y=E[\mathbf{y}_1\mathbf{y}_1^H]=\sum_{k=1}^K\sigma_s^2\mathbf{G}_k\mathbf{G}_k^H+\sigma_z^2\mathbf{I}_N$ is the covariance matrix of the received signal $\mathbf{y}_1$ and its unbiased sample covariance matrix can be simply estimated using the received signal samples [33].

As all the elements of $\mathbf{W}_R^{(j+1)}$ should have the same amplitude (that is, $\mathbf{W}_R^{(j+1)}\in\mathbf{Q}_R$), we update $\mathbf{W}_R^{(j+1)}$ from $\overline{\mathbf{W}}_R^{\text{opt}}$ as follows:

$$\mathbf{W}_R^{(j+1)}=\frac{1}{\sqrt{N_{\text{r}}}}\angle\overline{\mathbf{W}}_R^{\text{opt}}. \tag{10}$$

Similarly, for a given RF combining matrix $\mathbf{W}_R^{(j+1)}$, we update the digital baseband combining matrix to $\mathbf{W}_B^{(j+1)}$ by solving the unconstrained convex optimization problem

$$\min_{\mathbf{W}_B} E\left[\left\|\mathbf{s}_1-\mathbf{W}_B^H\mathbf{W}_R^{(j+1)H}\mathbf{y}_1\right\|^2\right]. \tag{11}$$

The optimal solution for $\mathbf{W}_B^{(j+1)}$ is given by

$$\mathbf{W}_B^{(j+1)}=\sigma_s^2\left(\mathbf{W}_R^{(j+1)H}\mathbf{R}_y\mathbf{W}_R^{(j+1)}\right)^{-1}\mathbf{W}_R^{(j+1)H}\mathbf{G}_1. \tag{12}$$

A step-by-step summary of the proposed hybrid combiner design is presented in Algorithm 1.

We define $\mathbf{W}=\mathbf{W}_R\mathbf{W}_B=[\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_{N_s}]$ with $\mathbf{w}_n\in C^{N_t\times 1}$, and $\mathbf{G}_k=[\mathbf{g}_{k,1}\ \mathbf{g}_{k,2}\ \ldots\ \mathbf{g}_{k,N_s}]$ with $\mathbf{g}_{k,n}\in C^{N_r\times 1}$. The signal-to-interference-plus-noise ratio (SINR) for the n-th symbol $s_{1,n}$, $n=1,2,\ldots,N_s$, of the desired UE is

Algorithm 1 Proposed hybrid combiner design based on alternating optimization

Require: $\mathbf{W}_{\text{R}}^{(0)}$ and $J_{\max}$

1: $\mathbf{W}_{\text{B}}^{(0)}=\sigma_s^2(\mathbf{W}_{\text{R}}^{(0)\text{H}}\mathbf{R}_y\mathbf{W}_{\text{R}}^{(0)})^{-1}\mathbf{W}_{\text{R}}^{(0)\text{H}}\mathbf{G}_1$

2: $j=0$

3: **while** $j\le J_{\max}$ **do**

4: $\overline{\mathbf{W}}_{\text{R}}^{\text{opt}}=\sigma_s^2\mathbf{R}_y^{-1}\mathbf{G}_1(\mathbf{W}_{\text{B}}^{(j)\text{H}}\mathbf{W}_{\text{B}}^{(j)})^{-1}\mathbf{W}_{\text{B}}^{(j)\text{H}}$

5: $\mathbf{W}_{\text{R}}^{(j+1)}=\frac{1}{\sqrt{N_{\text{r}}}}\text{R}\overline{\mathbf{W}}_{\text{R}}^{\text{opt}}$

6: $\mathbf{W}_{\text{B}}^{(j+1)}=\sigma_s^2(\mathbf{W}_{\text{R}}^{(j+1)\text{H}}\mathbf{R}_y\mathbf{W}_{\text{R}}^{(j+1)})^{-1}\mathbf{W}_{\text{R}}^{(j+1)\text{H}}\mathbf{G}_1$

7: $j=j+1$

8: **end while**

9: **Return** $\mathbf{W}_{\text{R}}^{(j)}$, $\mathbf{W}_{\text{B}}^{(j)}$

$$\text{SINR}_n=\frac{|\mathbf{w}_n^H\mathbf{g}_{1,n}|^2}{\sum\limits_{\substack{l=1\\l\neq n}}^{N_s}|\mathbf{w}_n^H\mathbf{g}_{1,l}|^2+\sum\limits_{k=2}^{K}\sum\limits_{l=1}^{N_s}|\mathbf{w}_n^H\mathbf{g}_{k,l}|^2+\frac{\sigma_z^2}{\sigma_s^2}||\mathbf{w}_n||^2}. \tag{13}$$

The achievable rate for the desired UE is given by

$$\Gamma=\sum_{n=1}^{N_s}\log_2(1+\text{SINR}_n). \tag{14}$$

4 | SIMULATION RESULTS

In this section, we compare the performance of the proposed hybrid combiner with those of the conventional schemes given in [28] and [30] and the unconstrained MMSE scheme, in terms of the MSE, achievable rate, and BER. The combining matrix $\mathbf{W}_{\text{o}}$ of the unconstrained MMSE scheme is obtained by solving the unconstrained convex optimization problem (6).

The simulation parameters are shown in Table 1. Here, a BS is equipped with $N_t=128$ transmit antennas and it serves $K=10$ UEs, each of which is assigned $N_s=2$ data

TABLE 1 Simulation parameters

Parameters	Values
Number of UEs, K	10
Number of transmit antennas at the BS, N_t	128
Number of receive antennas at the UE, N_r	4, 6
Number of RF chains at the UE, M_r	2
Number of data streams for the UE, N_s	2
Modulation order	16-QAM, 64-QAM

streams. The numbers of receive antennas and RF chains at the receiver of the desired UE are set as $N_r = 4, 6$ and $M_r = 2$, respectively. The modulation schemes for the information symbols are 16-ary quadrature amplitude modulation (16-QAM) and 64-QAM. Each element of the channel matrix $\mathbf{H}_1$ is generated by independent and identically distributed (i.i.d.) Rayleigh fading with zero mean and unit variance. The channel estimation error matrix $\mathbf{H}_{1,e}$ is generated by i.i.d. random variables with the distribution $CN(0, \sigma_e^2)$ and the CSI error variance is set to $\sigma_e^2 = 0.01$ [32]. The precoding matrix for each UE is designed as in Appendix, such that the precoding matrix for each UE is orthogonal to the estimated channel matrix $\hat{\mathbf{H}}_k$ of other UEs.

The unbiased sample covariance matrix is estimated by

$$\hat{\mathbf{R}}_y = \frac{1}{L-1} \sum_{l=1}^{L} \mathbf{y}_1(l)\mathbf{y}_1^H(l), \tag{15}$$

where $\mathbf{y}_1(l)$ indicates a received signal sample at UE 1 and L is the number of samples used to estimate the covariance matrix.

Figure 2 shows the MSE performances of the proposed scheme and the unconstrained MMSE scheme according to the parameter L when the sample covariance matrix is used, the number of receive antennas is $N_r = 4$, and the received SNR is 32 dB. In addition, the dotted green line represents the performance of the proposed scheme when the true covariance matrix is used. From the figure, we can observe that the performance of the proposed scheme with the estimated covariance matrix converges to that obtained with the true covariance matrix as the number of samples increases. For the rest of the simulation, we used $L = 4000$. The authors in [34] and [35] presented some schemes to reduce the number of samples used to estimate the covariance matrix while maintaining the estimation accuracy. However, this issue is out of the scope of the present study and it is not considered in this paper.

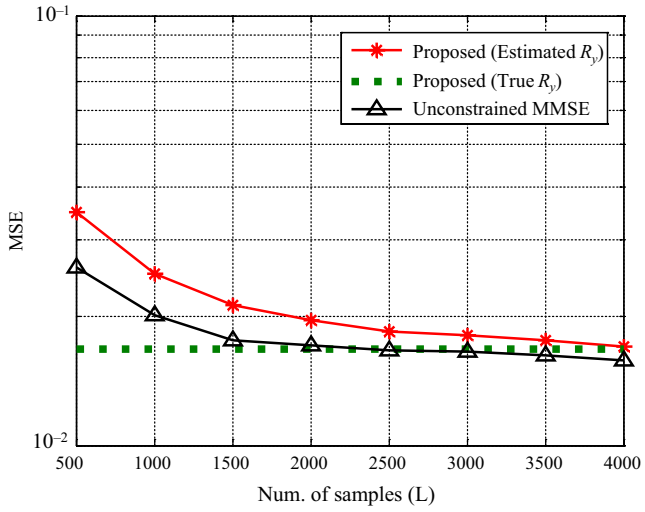

FIGURE 2 MSE comparison according to the number of samples used to estimate the covariance matrix

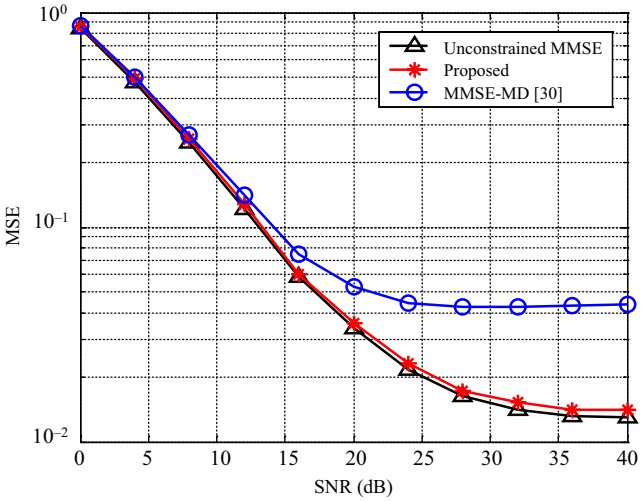

FIGURE 3 MSE comparison when $N_s = 2$ data streams are received through $N_r = 4$ receive antennas and $M_r = 2$ RF chains

Figure 3 shows the average MSEs of the proposed scheme, conventional MMSE-matrix decomposition (MMSE-MD) scheme in [30], and the unconstrained MMSE scheme when the number of receive antennas is $N_r = 4$ and the number of UEs is $K = 10$. In the conventional MMSE-MD scheme, the RF and baseband combining matrices are designed by applying the matrix decomposition to the unconstrained MMSE combining matrix. From the figure, we can observe that the proposed scheme has a lower MSE than the conventional MMSE-MD scheme. The performance difference between the proposed scheme and the MMSE-MD scheme increases as the SNR increases. This is because inter-user interference is a dominant factor in performance degradation in the high-SNR region, and the proposed scheme is better able to eliminate inter-user interference. Furthermore, we can observe that the conventional MMSE-MD scheme does not guarantee that the hybrid combining matrices $(\mathbf{W}_R, \mathbf{W}_B)$ obtained using the matrix decomposition method minimize the MSE given by $E[\| \mathbf{s}_1 - \mathbf{W}_o^H \mathbf{y}_1 \|^2]$, although the multiplication of the combining matrices $\mathbf{W}_R$ and $\mathbf{W}_B$ in the MMSE-MD scheme might be the closest to the unconstrained MMSE combining matrix $\mathbf{W}_o$ in terms of the Frobenius distance $\| \mathbf{W}_o - \mathbf{W}_R \mathbf{W}_B \|_F^2$, especially when there is residual inter-user interference at the receiver because of imperfect CSIT. Moreover, the performance of the proposed scheme is remarkably close to that of the unconstrained MMSE scheme even in the high-SNR region.

Figures 4 and 5 compare the achievable rate of the proposed scheme with those of the conventional and unconstrained MMSE optimal schemes. In the conventional SVD-MD scheme in [30], the RF and baseband combining matrices were designed by applying the matrix decomposition to the unconstrained SVD combining matrix. In the AGH-B-SVD scheme in [28], the RF combining matrix was chosen to harvest the array gain and the baseband combining matrix was designed based on the SVD of the desired channel matrix. The number of receive antennas is $N_r = 4$ and 6 for

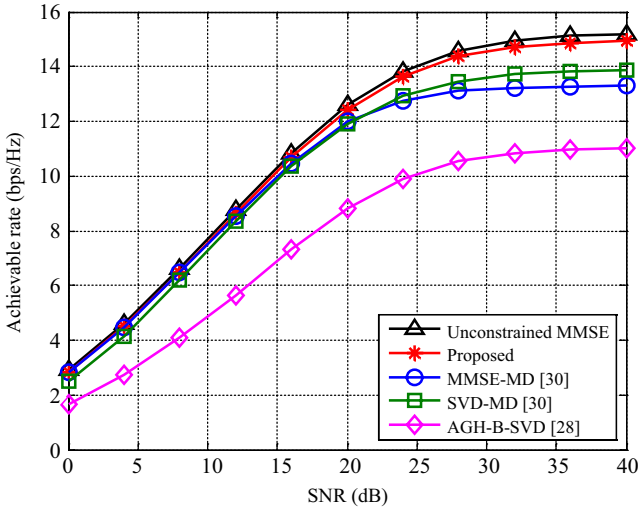

FIGURE 4 Achievable rate comparison when $N_s = 2$ data streams are received through $N_r = 4$ receive antennas and $M_r = 2$ RF chains

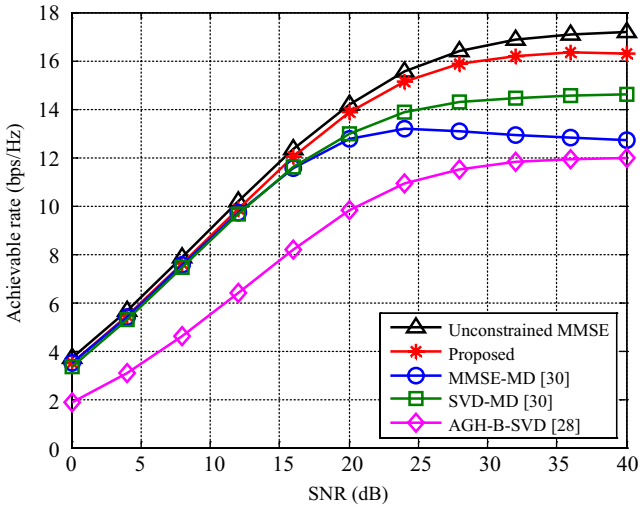

FIGURE 5 Achievable rate comparison when $N_s = 2$ data streams are received through $N_r = 6$ receive antennas and $M_r = 2$ RF chains

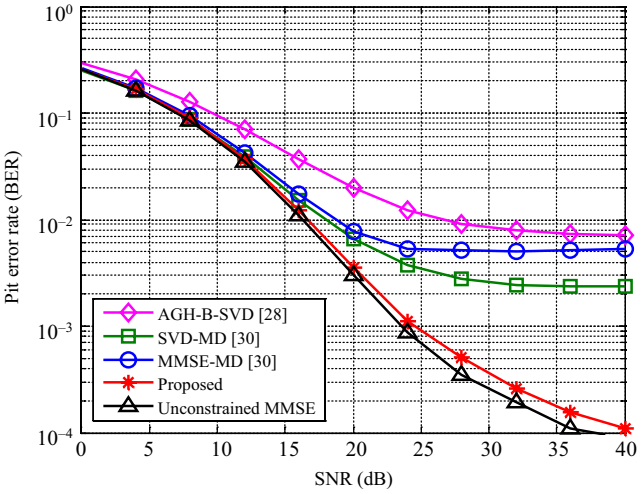

FIGURE 6 BER comparison of the UE with $N_r = 4$ when 16-QAM is used for modulation. The gray code is applied for the bit assignment of each symbol

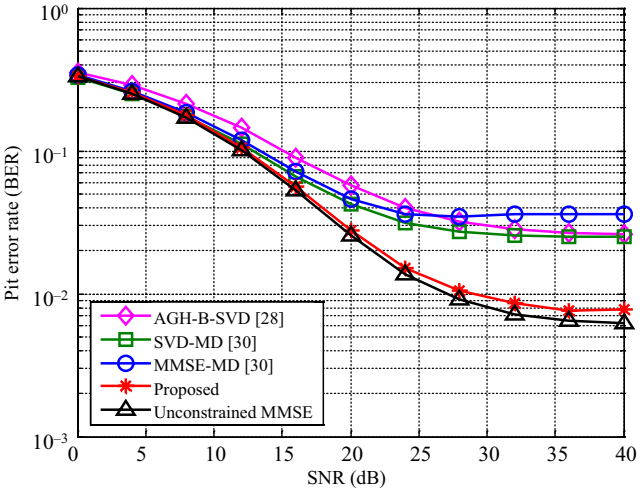

FIGURE 7 BER comparison of the UE with $N_r = 4$ when 64-QAM is used for modulation. The gray code is applied for the bit assignment of each symbol

Figures 4 and 5, respectively. By comparing the results in Figures 4 and 5, we can observe that the achievable rate increases as the number of receive antennas increases. This is because the SINR performance increases according to the number of receive antennas. Furthermore, it can be observed that the proposed scheme has a higher achievable rate than the conventional schemes for both $N_r = 4$ and 6. Similar to the MSE results in Figure 3, the achievable rate performance gap between the proposed scheme and the conventional schemes increases as the SNR increases.

Figures 6 and 7 show the BERs of the desired UE in the proposed scheme, the conventional schemes, and the unconstrained MMSE optimal scheme. The number of receive antennas in the desired UE is $N_r = 4$. In Figures 6 and 7, the modulation scheme for the information symbols is 16-QAM and 64-QAM, respectively. The gray code is applied for the bit assignment of each QAM symbol to minimize the BER. From the figures, it is observed that the proposed scheme achieves a BER performance similar to that of the unconstrained MMSE optimal scheme for both the 16-QAM and 64-QAM cases. However, the conventional schemes show extremely poor BER performance in the high-SNR region. The BER performance gap between the proposed scheme and the conventional schemes increases as the SNR increases for both cases (16-QAM and 64-QAM).

5 | CONCLUSION

In this paper, we presented a hybrid combiner design for downlink massive MU-MIMO systems, where each UE

has a limited number of RF chains—that is, less than the number of receive antennas—to reduce the implementation cost and power consumption. As it is exceedingly difficult to solve the original optimization problem owing to the constraint of constant amplitude on the RF combining matrix, we proposed an iterative alternating optimization scheme. In the proposed scheme, the analog RF and digital baseband combining matrices are alternatively optimized first without considering the constant amplitude constraint. This is accomplished by fixing one of them by turn and then adjusting the other matrix to satisfy the constant amplitude constraint. Using simulation results, we showed that the proposed scheme achieved better performance than the conventional schemes in the high-SNR region in terms of the MSE, achievable rate, and BER.

FUNDING INFORMATION

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministryof Education (2016R1D1A3B03935210).

ORCID

Bangwon Seo https://orcid.org/0000-0002-7152-1941

REFERENCES

1. D. Zhang et al., *Capacity analysis of NOMA with mmWave massive MIMO systems*, IEEE J. Sel. Areas Commun. **35** (2017), no. 7, 1606–1618.
2. Z. Ding et al., *NOMA meets finite resolution analog beamforming in massive MIMO and millimeter-wave networks*, IEEE Commun. Lett. **21** (2017), no. 8, 1879–1882.
3. L. N. Ribeiro et al., *Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs*, IEEE J. Sel. Topics Signal Process. **12** (2018), no. 3, 298–312.
4. G. Yang et al., *Low-latency heterogeneous networks with millimeter-wave communications*, IEEE Commun. Mag. **56** (2018), no. 6, 124–129.
5. S. M. R. Islam et al., *Resource allocation for downlink NOMA systems: Key techniques and open issues*, IEEE Wireless Commun. **25** (2018), no. 2, 40–47.
6. Y. Huang et al., *Signal processing for MIMO-NOMA: Present and future challenges*, IEEE Wireless Commun. **25** (2018), no. 2, 32–38.
7. G. Gui et al., *Deep learning for an effective non-orthogonal multiple access scheme*, IEEE Trans. Veh. Technol. **67** (2018), no. 9, 8440–8450.
8. H. Huang et al., *Rate region analysis in a full-duplex-aided cooperative nonorthogonal multiple-access system*, IEEE Access **5** (2017), 17869–17880.
9. T. L. Marzetta, *Noncooperative cellular wireless with unlimited numbers of base station antennas*, IEEE Trans. Wireless Commun. **9** (2010), no. 11, 3590–3600.
10. F. Rusek et al., *Scaling up MIMO: Opportunities and challenges with very large array*, IEEE Signal Process. Mag. **30** (2013), no. 1, 40–60.
11. Y.-H. Nam et al., *Full-dimension MIMO (FD-MIMO) for next generation cellular technology*, IEEE Commun. Mag. **51** (2013), no. 6, 172–179.
12. E. G. Larsson et al., *Massive MIMO for next generation wireless systems*, IEEE Commun. Mag. **52** (2014), no. 2, 186–195.
13. J. G. Andrews et al., *What will 5G be?* IEEE J. Sel. Areas Commun. **32** (2014), no. 6, 1065–1082.
14. L. Lu et al., *An overview of massive MIMO: Benefits and challenges*, IEEE J. Select. Topics Signal Process. **8** (2014), no. 5, 742–758.
15. A. Gupta and R. K. Jha, *A survey of 5G network: Architecture and emerging technologies*, IEEE Access **3** (2015), 1206–1232.
16. E. Bjornson, E. G. Larsson, and T. L. Marzetta, *Massive MIMO: ten myths and one critical question*, IEEE Commun. Mag. **54** (2016), no. 2, 114–123.
17. D. C. Araujo et al., *Massive MIMO: Survey and future research topics*, IET Commun. **10** (2016), no. 15, 1938–1946.
18. J. Zhang et al., *3D MIMO for 5G NR: Several observations from 32 to massive 256 antennas based on channel measurement*, IEEE Commun. Mag. **56** (2018), no. 3, 62–70.
19. H. Huang et al., *Deep learning for super-resolution channel estimation and DOA estimation based massive, MIMO system*, IEEE Trans. Veh. Technol. **67** (2018), no. 9, 8549–8560.
20. C.-S. Lee and W. H. Chung, *Hybrid RF-based precoding for cooperative multiuser massive MIMO systems with limited RF chains*, IEEE Trans. Commun. **65** (2014), no. 4, 1575–1589.
21. A. Garcia-Rodriguez et al., *Hybrid analog-digital precoding revisited under realistic RF modeling*, IEEE Wireless Commun. Lett. **5** (2016), no. 5, 528–531.
22. N. Li et al., *Hybrid precoding for mmWave massive MIMO systems with partially connected structure*, IEEE Access **5** (2017), 15142–15151.
23. S. Park et al., *Exploiting spatial channel covariance for hybrid precoding in massive MIMO systems*, IEEE Trans. Signal Process. **65** (2017), no. 14, 3818–3832.
24. T.-Y. Chang and C.-E. Chen, *A hybrid Tomlinson-Harashima transceiver design for multiuser mmWave MIMO systems*, IEEE Wireless Commun. Lett. **7** (2018), no. 1, 118–121.
25. M. Li et al., *Joint hybrid precoder and combiner design for multistream transmission in mmWave MIMO systems*, IET Commun. **11** (2017), no. 17, 2596–2604.
26. Y. Feng and S. C. Kim, *Hybrid precoding in point-to-point massive multiple-input multiple-output systems based on normalized matrix adaptive method*, IET Commun. **11** (2017), no. 12, 1882–1885.
27. A. Liu and V. Lau, *Phase only RF precoding for massive MIMO systems with limited RF chains*, IEEE Trans. Signal Process. **62** (2014), no. 17, 4505–4515.
28. W. Ni and X. Dong, *Hybrid block diagonalization for massive multiuser MIMO systems*, IEEE Trans. Commun. **64** (2016), no. 1, 201–211.
29. R. Rajashekar and L. Hanzo, *Iterative matrix decomposition aided block diagonalization for mm-wave multiuser MIMO systems*, IEEE Trans. Wireless Comm. **16** (2017), no. 3, 1372–1384.
30. W. Ni, X. Dong, and W.-S. Lu, *Near-optimal hybrid processing for massive MIMO systems via matrix decomposition*, IEEE Trans. Sig. Process. **65** (2017), no. 15, 3922–3933.

31. L. Zhao, D. W. K. Ng, and J. Yuan, *Multi-user precoding and channels estimation for hybrid millimeter wave systems*, IEEE J. Sel. Areas Commun. **35** (2017), no. 7, 1576–1590.
32. N. Kim, Y. Lee, and H. Park, *Performance analysis of MIMO system with linear MMSE receiver*, IEEE Trans. Wireless Commun. **7** (2008), no. 11, 4474–4478.
33. K. V. Mardia, J. T. Kent, and J. M. Bibby, Multivariate analysis (probability and mathematical statistics), Academic Press, San Diego, 1995.
34. A. J. Rothman et al., *Sparse permutation invariant covariance estimation*, Electron. J. Stat. **2** (2008), 494–515.
35. M. Yuan, *High dimensional inverse covariance matrix estimation*, J. Mach. Learn. Res. **11** (2010), 2261–2286.

AUTHOR BIOGRAPHY

Bangwon Seo received his BS, MS, and PhD degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 1997, 1999, and 2010, respectively. From November 2004 to February 2013, he was with the Electronics Telecommunications Research Institute, Daejeon, Republic of Korea. In March 2013, he joined the Division of Electrical, Electronic and Control Engineering, Kongju National University, Cheonan, Republic of Korea, where he is currently an associate professor. His research interests include massive MIMO, hybrid transceiver, device-to-device direct communications, MIMO systems, and OFDM systems.

APPENDIX

In this study, the transmit precoding matrix is designed by following a similar approach as in [30]. First, a precoding matrix with full RF chains is designed without considering the constraint on the number of RF chains. To eliminate interference in (1), the full-RF-chain transmit precoding matrix $\mathbf{F}_k^o, k=1,...,K$ needs to be orthogonal to the estimated channel $\hat{\mathbf{H}}_j$ for all $j\neq k$. This orthogonality condition can be written as $\hat{\mathbf{H}}_j\mathbf{F}_k^o=\mathbf{0}_{N\mathrm{r}\times Ns}$, $j\neq k$. Now, we define the interference channel for UE k as

$$\overline{\mathbf{H}}_k=[\,\hat{\mathbf{H}}_1^T,...,\hat{\mathbf{H}}_{k-1}^T\,\hat{\mathbf{H}}_{k+1}^T,...,\hat{\mathbf{H}}_K^T\,]^T. \tag{A1}$$

Then, the orthogonality condition can be rewritten as $\overline{\mathbf{H}}_k\mathbf{F}_k^o=\mathbf{0}_{(K-1)N\mathrm{r}\times Ns}$. If we assume that $N_t\geq(K-1)N_\mathrm{r}+N_\mathrm{s}$, $\mathbf{F}_k^o$ can be selected as a subset of the null-space basis of $\overline{\mathbf{H}}_k$.

More specifically, the SVD of $\overline{\mathbf{H}}_k$ is given by

$$\overline{\mathbf{H}}_k=\mathbf{U}_k\Sigma_k\mathbf{V}_k^H, \tag{A2}$$

where $\mathbf{U}_k$ and $\mathbf{V}_k$ are unitary matrices of sizes $(K-1)N_\mathrm{r}\times(K-1)N_\mathrm{r}$ and $N_\mathrm{t}\times N_\mathrm{t}$, respectively. $\sum_k$ is a rectangular diagonal matrix of size $(K-1)N_\mathrm{r}\times N_\mathrm{t}$ with r_k nonnegative singular values on the diagonal where r_k is the rank of $\overline{\mathbf{H}}_k$ and $r_k\leq(K-1)N_\mathrm{r}$. The matrix $\mathbf{V}_k$ can be partitioned as $\mathbf{V}_k=[\mathbf{V}_{k,1}\,\mathbf{V}_{k,2}]$ where the sizes of the matrices $\mathbf{V}_{k,1}$ and $\mathbf{V}_{k,2}$ are $N_\mathrm{t}\times r_k$ and $N_\mathrm{t}\times(N_\mathrm{t}-r_k)$, respectively, and $\mathbf{V}_{k,2}$ is a null-space basis of $\overline{\mathbf{H}}_k$. Hence, to satisfy the orthogonality condition, $\mathbf{F}_k^o$ can be selected as a subset of $\mathbf{V}_{k,2}$.

Now, let us consider the design of a hybrid precoding matrix by considering the constraint on the number of RF chains. If we denote the number of RF chains at the BS as M_t, the hybrid precoding matrix can be written as $\mathbf{F}_k=\mathbf{F}_{k,R}\mathbf{F}_{k,B}$ where $\mathbf{F}_{k,R}$ and $\mathbf{F}_{k,B}$ represent an analog RF precoding matrix of size $N_\mathrm{t}\times M_\mathrm{t}$ and a digital baseband precoding matrix of size $M_\mathrm{t}\times N_\mathrm{s}$, respectively. As $\mathbf{F}_{k,R}$ is implemented by analog phase shifters, all the elements of $\mathbf{F}_{k,R}$ are constrained to have a constant amplitude such that $|[\mathbf{F}_{k,R}]_{m,n}|=1/\sqrt{N_\mathrm{t}}$.

Using a similar approach as in [30], the hybrid precoding matrices $(\mathbf{F}_{k,R},\mathbf{F}_{k,B})$ are designed to minimize the Frobenius distance between $\mathbf{F}_k^o$ and $\mathbf{F}_{k,R}\mathbf{F}_{k,B}$ given by

$$\min_{\mathbf{F}_{k,R},\mathbf{F}_{k,B}} \|\,\mathbf{F}_k^o-\mathbf{F}_{k,R}\mathbf{F}_{k,B}\,\|_F^2 \ \text{s.t.}\mathbf{F}_{k,R}\in\mathrm{Q}_R. \tag{A3}$$

The solution of this optimization problem is obtained using the iterative alternating optimization method given in [Algorithm 1, 30]. The authors in [30] showed that the performance of this hybrid precoding scheme approaches that of the full-RF-chain precoding scheme when there is no channel estimation error.