

Hybrid Precoding and Combining Design for Millimeter-Wave Multi-User MIMO based on SVD

Ang Li and Christos Masouros

Dept. of Electronic and Electrical Eng., University College London, London, U.K.

Email: {ang.li.14, c.masouros}@ucl.ac.uk

Abstract—In this paper, we focus on the millimeter-wave multi-user multiple-input-multiple-output (mmWave MU-MIMO) systems and propose a low-complexity hybrid precoding and combining design, which is applicable to both fully-connected structures and sub-connected structures. Based on the channel knowledge of each user, the analog combiner for each user is independently designed based on the singular value decomposition (SVD), while the analog precoder is obtained by the conjugate transposition to maximize the effective channel gain. Then, with the resulting effective analog channel, low-dimensional baseband precoders can be efficiently applied. The proposed scheme requires no optimization techniques or any complicated iterative algorithms, while the numerical results show that it can approach the performance of fully digital schemes and even achieve a better performance in some scenarios. It is also observed that sub-connected structures can achieve a much higher power efficiency compared to fully-connected structures and are therefore promising for the future green communication systems.

Index Terms—Millimeter-Wave, MU-MIMO, hybrid design, SVD, power efficiency.

I. INTRODUCTION

With the rich and under-exploited spectrum resources in the millimeter-wave (mmWave) band, mmWave communications will be promising for the future fifth-generation (5G) cellular systems [1]-[3]. Compared to existing microwave cellular systems, the operating frequency for mmWave communications is in GHz, which results in a much higher pathloss in the free-space [4]-[5]. Then, to overcome the increasing pathloss and offer satisfactory signal-to-noise-ratio (SNR), mmWave communication systems will require a large number of antennas at the base station (BS) [2]. Thanks to the high frequency, the small carrier wavelength for mmWave enables packing hundreds of antennas at the BS. For a very large scale MIMO system, theoretical studies have shown that simple linear precoding schemes can achieve near-optimal performance [6]-[7]. Nevertheless, the conventional fully digital precoding techniques require a dedicated radio frequency (RF) chain for each antenna element, which has difficulty for practical implementation at present due to the high cost and power consumption, especially for mixed-signal hardware designed for mmWave band.

Due to the practical constraints above, initial designs for current mmWave communications have focused on analog only processing with low-cost phase shifters [8]-[9]. However, there exists a performance gap between the analog only designs and the fully digital schemes. Furthermore, RF process-

ing schemes have limited capability in managing interference, and are therefore not promising for multi-user communication systems. To enable spatial multiplexing and approach the performance of fully digital schemes, hybrid designs have been further proposed in [10]-[14], where a reduced number of RF chains is required, and the pre-processing has been divided into the analog and digital domain. In the analog domain, phase shifters are applied to control the phases to harvest the array gain, while in the digital domain, a low-dimensional baseband precoder is applied to manage the multi-stream or multi-user interference. Another advantage of the hybrid designs is the increase in the power efficiency thanks to the reduced number of RF chains. In [10]-[11], hybrid designs for single-user mmWave MIMO systems have been proposed, where in [10] an orthogonal match pursuit (OMP) based approach is proposed to obtain the analog and digital solution, and an iterative hybrid design is further proposed in [11]. The extensions to multi-user case have been introduced in [12]-[14], where in [12] a codebook based scheme is proposed to obtain the analog precoder and combiner, and then the zero-forcing (ZF) precoder is applied in the baseband. A modified scheme has been included in [13] where a minimum-mean-squared-error (MMSE) approach is employed upon the OMP approach to obtain further capacity improvements for low SNR scenarios.

We note that the schemes designed in [10]-[14] all assume a fully-connected structure at the transmit antenna array, where each antenna element is connected to all RF chains. However, the complicated fully-connected structure requires a large number of switches and phase shifters, which will introduce significant insertion losses in practice [15] and the insertion losses have an inferior effect on the system performance. Based on the above, sub-connected structures where each antenna element is connected to only one RF chain are introduced and shown to be promising for practical implementation due to both the simplified antenna structures and the reduced amount of insertion losses. While the above designs cannot be trivially extended to sub-connected structures, a hybrid scheme that considers the sub-connected structures has been proposed in [16] based on the alternating minimization (AltMin) algorithms. A successive interference cancellation (SIC)-based approach designed for sub-connected structures is also proposed in [17]. However, these two schemes are designed for the single-user case and not trivially extended to multi-user case.

In this paper, a hybrid precoder and combiner design for

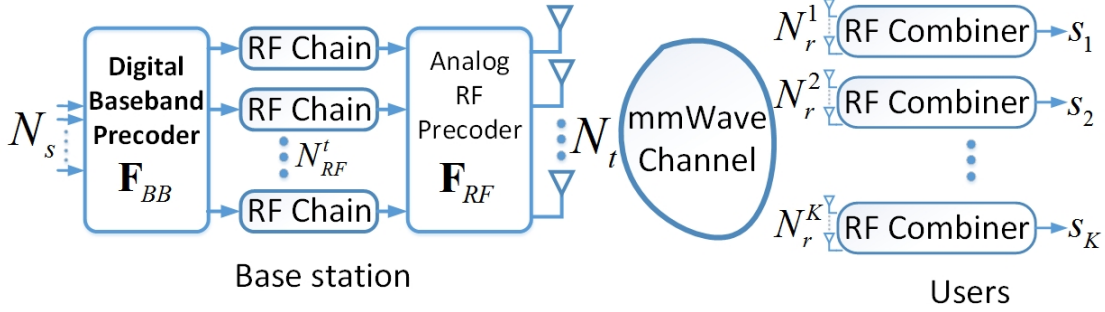


Fig. 1: Block diagram of the system model

mmWave MU-MIMO systems is proposed. For each user, the analog only combiner is designed based on the SVD of the user's channel, and the analog precoder at the base station (BS) is obtained based on the channel conjugate transposition to maximize the effective channel gain. Then, the low-dimensional baseband precoder is obtained based on the effective analog channel. We note that only the knowledge of the channel is needed for the proposed scheme, while no further feedback is necessary between the BS and each user. Compared with existing approaches, the proposed scheme has the following advantages: 1) the complexity is low and it requires no optimization methods or any complicated iterative designs; 2) it can achieve a very competitive performance in various system settings; 3) it can be trivially extended to sub-connected structures. Based on the above advantages, the proposed scheme is therefore promising for practical implementation. We further compare the power efficiency performance between fully-connected and sub-connected structures, where it is observed that sub-connected structures are much more energy efficient due to the reduced number of phase shifters, which makes them more favourable for the future communication systems that target at green communications.

Notations: a , \mathbf{a} , and \mathbf{A} denote scalar, vector and matrix, respectively. $\mathbb{E}\{\cdot\}$, $(\cdot)^H$, and $(\cdot)^{-1}$ denote expectation, conjugate transpose, and inverse of a matrix respectively. $\|\cdot\|_F$ denotes the Frobenius norm, \mathbf{I} is the identity matrix and $\mathbf{0}$ denotes a zero matrix or vector. $\mathcal{C}^{n \times n}$ represents a $n \times n$ matrix in the complex set. $diag(\cdot)$ denotes the conversion of a vector into a diagonal matrix with the values on its main diagonal. $\mathbf{R}(k, u)$ denote the element of the k th row and u th column in \mathbf{R} .

II. SYSTEM DESCRIPTION

A. System and Signal Model

A downlink mmWave MU-MIMO system is considered, where a BS with N_t transmit antennas and N_{RF}^t RF chains is assumed. A block diagram of the system is given in Fig. 1, and the transmit array structure is shown in Fig. 2, where the number of antennas connected to each RF chain for sub-connected structures is $M = N_t/N_{RF}^t$. There is a total number of K mobile stations (MSs) in the system, and each MS k is equipped with N_r^k antennas. Following [12]-[14], we focus on the case where the BS employs K RF chains to

support the transmission of K streams, i.e. $K = N_{RF}^t$. Perfect knowledge of the channel state information (CSI) is assumed at the BS, which is a common assumption in precoding designs for mmWave MIMO systems [13][14]. At the BS side, a hybrid precoder is considered, where the precoder is decomposed into the analog precoder and baseband precoder. Then, we can express the transmit signal vector as

$$\mathbf{x} = \mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s}, \quad (1)$$

where we denote

$$\mathbf{F}_{BB} = [\mathbf{f}_{BB1}, \mathbf{f}_{BB2}, \dots, \mathbf{f}_{BBK}] \in \mathcal{C}^{K \times K} \quad (2)$$

as the baseband precoder. $\mathbf{s} \in \mathcal{C}^{K \times 1}$ is denoted as the data symbol vector and for simplicity $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \frac{P}{K} \cdot \mathbf{I}$, where P is the average total transmit power and we assume equal power allocation for each stream. $\mathbf{F}_{RF} \in \mathcal{C}^{N_t \times K}$ is the analog precoder implemented with phase shifters, and therefore each entry of \mathbf{F}_{RF} is of constant modulus. When a fully-connected structure is considered as shown in Fig. 2 (a), \mathbf{F}_{RF} can be expressed as

$$\mathbf{F}_{RF} = [\mathbf{f}_1^{full}, \mathbf{f}_2^{full}, \dots, \mathbf{f}_K^{full}], \quad (3)$$

where each $\mathbf{f}_k^{full} \in \mathcal{C}^{N_t \times 1}$. In this case, \mathbf{F}_{RF} is a full matrix with all elements being non-zero and normalized to satisfy

$$|\mathbf{f}_k^{full}(m)|^2 = \frac{1}{N_t}. \quad (4)$$

For sub-connected structures as in Fig. 2 (b), \mathbf{F}_{RF} becomes a block diagonal matrix and can be expressed as

$$\mathbf{F}_{RF} = diag(\mathbf{f}_1^{sub}, \mathbf{f}_2^{sub}, \dots, \mathbf{f}_K^{sub}), \quad (5)$$

where each $\mathbf{f}_k^{sub} \in \mathcal{C}^{M \times 1}$. In this case, each entry of \mathbf{f}_k^{sub} satisfies

$$|\mathbf{f}_k^{sub}(m)|^2 = \frac{1}{M}. \quad (6)$$

The total power constraint is satisfied by normalizing \mathbf{F}_{BB} such that $\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 = K$. At the receiver side, each MS k is equipped with one RF chain, and a RF only combiner $\mathbf{w}_k^H \in \mathcal{C}^{1 \times N_r^k}$ is applied, where each entry of the vector \mathbf{w}_k satisfies

$$|\mathbf{w}_k(m)|^2 = \frac{1}{N_r^k}. \quad (7)$$

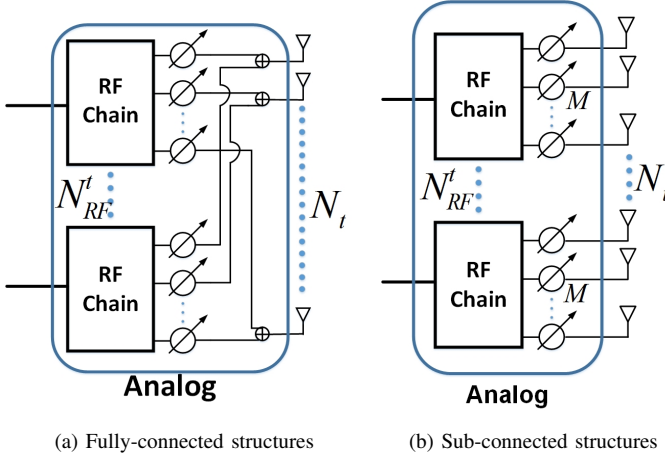


Fig. 2: Transmit array structures

Then, the received symbol for MS k can be expressed as

$$r_k = \mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{f}_{BBk} s_k + \mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{RF} \sum_{j \neq k} \mathbf{f}_{BBj} s_j + \mathbf{w}_k^H \mathbf{n}_k, \quad (8)$$

where $\mathbf{H}_k \in \mathcal{C}^{N_r^k \times N_t}$ denotes the mmWave channel from the BS to MS k , \mathbf{n}_k is the circular symmetric Gaussian distributed additive noise vector with zero mean and covariance $\sigma^2 \cdot \mathbf{I}$. Due to the unitary transformation, $\mathbf{w}_k^H \mathbf{n}_k$ has the same distribution as \mathbf{n}_k . Then, the received signal-to-interference-plus-noise-ratio (SINR) for the k th user can be obtained as

$$\gamma_k = \frac{\frac{P}{K} \|\mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{f}_{BBk}\|_F^2}{\sum_{j \neq k} \frac{P}{K} \|\mathbf{w}_k^H \mathbf{H}_k \mathbf{F}_{RF} \mathbf{f}_{BBj}\|_F^2 + \sigma^2}. \quad (9)$$

By assuming Gaussian input, the achievable sum rate of the system can be obtained as

$$R = \sum_{k=1}^K \log_2(1 + \gamma_k). \quad (10)$$

B. Millimeter-Wave MIMO Channel Model

MmWave channels are expected to have limited scattering [2][5]. To take this effect into consideration, a geometric channel model with L_u independent propagation paths is considered for each \mathbf{H}_k [12][13], where the value of L_u is small compared to N_t for limited scattering. Then, each \mathbf{H}_k can be expressed as

$$\mathbf{H}_k = \sqrt{\frac{N_t N_r^k}{L_u}} \sum_{l=1}^{L_u} \alpha_{k,l} \cdot \mathbf{a}_{MS}(\theta_{k,l}^{MS}, \phi_{k,l}^{MS}) \mathbf{a}_{BS}^H(\theta_{k,l}^{BS}, \phi_{k,l}^{BS}), \quad (11)$$

where $\alpha_{k,l}$ is the complex path gain and is assumed to follow i.i.d. standard complex Gaussian distribution that forms the Rayleigh component. $\mathbf{a}_{MS}(\theta_{k,l}^{MS}, \phi_{k,l}^{MS})$ and $\mathbf{a}_{BS}^H(\theta_{k,l}^{BS}, \phi_{k,l}^{BS})$ are receive and transmit array response vectors respectively, where $\theta_{k,l}^{MS}$ ($\phi_{k,l}^{MS}$) and $\theta_{k,l}^{BS}$ ($\phi_{k,l}^{BS}$) denote the azimuth (elevation) angles of arrival and departure (AoAs/AoDs) of the l th path. For uniform linear arrays

(ULAs), only the azimuth AoAs and AoDs are considered and we can express the array response as

$$\mathbf{a}_{ULA}(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{jkd \sin(\theta)}, \dots, e^{j(N-1)kd \sin(\theta)} \right]^T, \quad (12)$$

where $k = \frac{2\pi}{\lambda}$ with λ being the carrier wavelength, and d is the inter-antenna spacing. When uniform planar arrays (UPAs) are considered, the array response can be expressed as

$$\mathbf{a}_{UPA}(\theta, \phi) = \frac{1}{\sqrt{N}} \left[1, \dots, e^{jkd(m \sin(\phi) \sin(\theta) + n \cos(\theta))}, \dots, e^{jkd((W-1) \sin(\phi) \sin(\theta) + (H-1) \cos(\theta))} \right]^T, \quad (13)$$

where $0 \leq m \leq W-1$, $0 \leq n \leq H-1$, and $N = WH$. $N = N_t$ for \mathbf{a}_{BS} and $N = N_r^k$ for \mathbf{a}_{MS} of \mathbf{H}_k . While we consider ULAs and UPAs in the simulations, the proposed schemes apply to any other array structures.

III. PROPOSED SCHEME BASED ON SVD

We present the proposed scheme in this section, where the phase-only control is applied for the analog precoder and combiner. The analog combiner \mathbf{w}_k for each MS k is designed based on the knowledge of \mathbf{H}_k , and then we design the analog precoder \mathbf{F}_{RF} at the BS. We firstly propose the hybrid scheme for the fully-connected antenna array, and then extend to sub-connected structures. By denoting $\mathbf{W}_{RF} = \text{diag}[\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K]$ and $\mathbf{H} = [\mathbf{H}_1^T, \mathbf{H}_2^T, \dots, \mathbf{H}_K^T]^T$, we can express the effective analog channel in a compact form as

$$\mathbf{H}_{eff} = \mathbf{W}_{RF}^H \mathbf{H} \mathbf{F}_{RF}. \quad (14)$$

Then, based on the effective analog channel, a low-dimensional baseband processing can be applied to handle the multi-user interference. We note that the diagonal entries of \mathbf{H}_{eff} denote the effective analog channel gains, while the off-diagonal entries denote analog multi-stream interference. Therefore for the designs of the analog precoder and combiner, we focus on the diagonal entries of \mathbf{H}_{eff} and propose to maximize the effective analog channel gains for each user, while the off-diagonal interference can be well managed by the baseband digital precoder \mathbf{F}_{BB} . For fully-connected structures, the k -th diagonal term of \mathbf{H}_{eff} can be expressed as

$$\mathbf{H}_{eff}(k, k) = \mathbf{w}_k^H \mathbf{H}_k \mathbf{f}_k^{full}. \quad (15)$$

To reduce the complexity and avoid further information exchange between the BS and MSs, we propose to design the combiners and the precoders separately, where each combiner \mathbf{w}_k is obtained by the SVD of the channel \mathbf{H}_k , while the analog precoder \mathbf{F}_{RF} is obtained by the conjugate transposition. The detailed designs are introduced below.

A. Analog Combiner Design

For each MS k , we firstly express the SVD of the channel matrix \mathbf{H}_k as

$$\mathbf{H}_k = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^H, \quad (16)$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{N_r^k}]$ and \mathbf{V} are unitary matrices that contain the left and right singular vectors. $\mathbf{\Lambda}$ is a diagonal

matrix with the singular values on the diagonal in a descending order. For sum rate maximization, \mathbf{w}_k is then selected as the normalized singular vector which corresponds to the largest singular value, expressed as

$$\mathbf{w}_k(m) = \frac{1}{\sqrt{N_r^k}} e^{j \cdot \varphi_m}, \quad m \in \{1, 2, \dots, N_r^k\}, \quad (17)$$

where φ_m is the phase of the m th element in \mathbf{u}_1 . As the design of \mathbf{w}_k only requires the knowledge of \mathbf{H}_k , each MS k can obtain \mathbf{w}_k individually.

It is known that the SVD involves high computational complexity and is difficult to implement in hardware. To reduce the complexity, we note that each \mathbf{u}_i is also the eigenvector for $\mathbf{H}_k \mathbf{H}_k^H$ and we only need to obtain the eigenvector \mathbf{u}_1 that corresponds to the largest eigenvalue. Therefore, in practice we can employ the ‘‘power iteration’’ algorithm in [18], which is employed to compute the largest eigenvalue and the corresponding eigenvector of a diagonalizable matrix. Since $\mathbf{H}_k \mathbf{H}_k^H$ is Hermitian and diagonalizable, in this way we can utilize the power iteration algorithm to efficiently compute \mathbf{u}_1 and the complicated SVD decomposition can be avoided.

B. Analog Precoder Design

After we obtain \mathbf{w}_k , for each MS k we denote

$$\mathbf{t}_k = \mathbf{w}_k^H \mathbf{H}_k. \quad (18)$$

Recall (15) and we note that $\mathbf{t}_k \in \mathcal{C}^{1 \times N_t}$, and therefore to maximize the effective channel gain $\mathbf{H}_{eff}(k, k)$, \mathbf{f}_k^{full} for each user k is selected as the conjugate transposition of \mathbf{t}_k with an element-wise normalization, expressed as

$$\mathbf{f}_k^{full}(m) = \frac{1}{\sqrt{N_t}} e^{j \cdot (-\omega_m)}, \quad m \in \{1, 2, \dots, N_t\}, \quad (19)$$

where ω_m is the phase of the m th element in the vector \mathbf{t}_k and $-\omega_m$ represents the conjugate transposition. This precoder is similar to the matched filter but the normalization here is applied for each element in $\mathbf{f}_k^{full}(m)$ due to the constant modulus constraint of the phase shifters. Based on (18)(19), we observe that the design of the analog precoder requires the knowledge of the corresponding analog combiner. Therefore, at the BS, for each user k the BS firstly obtains the eigenvector \mathbf{u}_1 that corresponds to the largest eigenvalue of \mathbf{H}_k via ‘‘power iteration’’ scheme in [18] based on the CSI, and then calculates each \mathbf{w}_k based on (17). With the analog combiner \mathbf{w}_k and further \mathbf{t}_k obtained at the BS for each user k , \mathbf{F}_{RF} can then be efficiently calculated based on (19).

We note that while a joint analog combiner and precoder design can be applied to jointly select the normalized eigenvectors and analog precoders that maximize the effective analog channel gains, the case that the optimal \mathbf{w}_k is selected as the normalized eigenvector that corresponds to the largest eigenvalue is with probability one. Therefore, instead of applying a joint approach that requires additional information exchange and higher complexity, the proposed decomposed scheme is more practical and more beneficial in terms of complexity.

C. Baseband Precoder Design

With each \mathbf{w}_k and \mathbf{F}_{RF} obtained by (17) and (19), the effective analog channel can be obtained by (14). Then at the baseband, for simplicity a closed-form ZF or MMSE precoder is applied to manage the multi-user interference, expressed as

$$\begin{aligned} \mathbf{F}_{BB}^{ZF} &= f_{ZF} \cdot \mathbf{H}_{eff}^H (\mathbf{H}_{eff} \mathbf{H}_{eff}^H)^{-1}, \\ \mathbf{F}_{BB}^{MMSE} &= f_{MMSE} \cdot \mathbf{H}_{eff}^H (\mathbf{H}_{eff} \mathbf{H}_{eff}^H + K \sigma^2 \cdot \mathbf{I})^{-1}. \end{aligned} \quad (20)$$

In (20), f_{ZF} and f_{MMSE} are the normalization factors to ensure the power constraint $\|\mathbf{F}_{RF} \mathbf{F}_{BB}\|_F^2 = K$, expressed as

$$\begin{aligned} f_{ZF} &= \frac{\sqrt{K}}{\left\| \mathbf{F}_{RF} \cdot \mathbf{H}_{eff}^H (\mathbf{H}_{eff} \mathbf{H}_{eff}^H)^{-1} \right\|_F}, \\ f_{MMSE} &= \frac{\sqrt{K}}{\left\| \mathbf{F}_{RF} \cdot \mathbf{H}_{eff}^H (\mathbf{H}_{eff} \mathbf{H}_{eff}^H + K \sigma^2 \cdot \mathbf{I})^{-1} \right\|_F}. \end{aligned} \quad (21)$$

Remark: While in this paper only ZF and MMSE approaches are considered in the baseband, with the proposed ‘effective analog channel’ approach any digital precoding schemes such as constructive interference or transmit beamforming designed for conventional MIMO can be efficiently applied for mmWave MIMO systems, as the effective channel is low-dimensional, which is another advantage of the proposed scheme.

D. Extension to Sub-connected structures

When sub-connected structures are considered, \mathbf{F}_{RF} is block-diagonal as given in (5), where each $\mathbf{f}_k^{sub} \in \mathcal{C}^{M \times 1}$. Similar to the design for the fully-connected structures where the conjugate transposition is applied, as shown in (19), \mathbf{f}_k^{sub} is designed as

$$\mathbf{f}_k^{sub}(m) = \frac{1}{\sqrt{M}} e^{j \cdot (-\omega_n)}, \quad m \in \{1, 2, \dots, M\}, \quad (22)$$

where ω_n is the phase of the n -th element in the vector \mathbf{t}_k . In (22), $n = (k-1)M + m$ and we have $KM = N_t$. As can be observed, compared to fully-connected structures where each $\mathbf{f}_k^{full} \in \mathcal{C}^{N_t \times 1}$, for sub-connected structures $\mathbf{f}_k^{sub} \in \mathcal{C}^{M \times 1}$ can only harvest part of the array gains, and therefore the sub-connected structures will achieve a lower sum rate compared to the fully-connected structures, which will be shown in the simulation results.

IV. POWER EFFICIENCY EVALUATION

Future communication systems target at green communications and therefore the power efficiency becomes an important performance metric. In this paper, we compare the power efficiency of the fully-connected and sub-connected structures for mmWave hybrid precoding schemes, where the power efficiency is expressed as

$$\eta = \frac{R}{P_{full/sub}}, \quad (23)$$

which is defined as the achievable sum rate per unit power consumption. In (23), $P_{fully/sub}$ denotes the total power

consumption for fully-connected or sub-connected structures. The power consumption model employed in this paper can be found in [19], where the power consumption for fully-connected and sub-connected structures can be respectively expressed as

$$\begin{aligned}
 P_{full} &= N_t (N_{RF}^t + 1) P_{PA} + N_t N_{RF}^t P_{PS} \\
 &\quad + N_{RF}^t (P_{RFC} + P_{DAC}) + P_{BB} \\
 P_{sub} &= N_t P_{PA} + N_t P_{PS} + N_{RF}^t (P_{RFC} + P_{DAC}) + P_{BB}.
 \end{aligned} \tag{24}$$

In (24), P_{PA} is the power for the power amplifiers, P_{PS} is the power for the phase shifters, P_{RFC} represents the power for the RF chains, P_{DAC} denotes the power for the digital-to-analog converters, and P_{BB} is the power for the baseband processing. It can be observed from (24) that the fully-connected structures consume more power on P_{PA} and P_{PS} , which is because the fully-connected structures require more switches and phase shifters. The values of each parameter employed in this paper are as follows: $P_{PA} = 20\text{mW}$, $P_{PS} = 30\text{mW}$, $P_{RFC} = 30\text{mW}$, $P_{DAC} = 200\text{mW}$, and $P_{BB} = 5\text{mW}$ [19].

V. NUMERICAL RESULTS

We present the numerical results in this section based on Monte Carlo simulations. We assume the number of paths for the mmWave channel is $L_u = 6$, and the antenna spacing is $d = \frac{\lambda}{2}$. The azimuth AoAs/AoDs are assumed to be uniformly distributed in $[0, 2\pi]$, and elevation AoAs/AoDs follow a uniform distribution in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ for UPAs. For each MS k , for simplicity we assume an identical number of receive antennas, i.e. $N_r^k = N_r^0, \forall k$. The proposed scheme is compared with the existing codebook based scheme in [12] and the unconstrained digital block diagonalization (BD) in [20]. The SNR in each plot is defined as $\rho = \frac{1}{K\sigma^2}$, where it is assumed that the total transmit power $P = 1$.

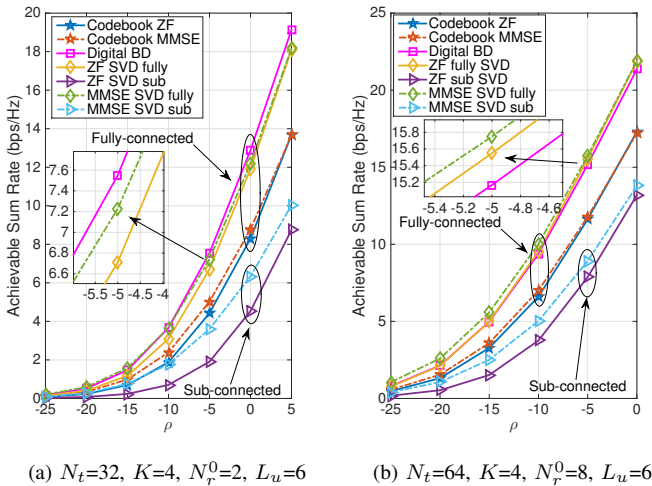


Fig. 3: Sum rate results for ULAs

In Fig. 3, the sum rate of different schemes with ULAs is shown, where there is a total number $K = 4$ MSs in the system. For the codebook based scheme, the codebook size is 5 bits. In both figures, a significant performance gain can be observed for the proposed schemes compared to the existing

schemes, and the performance is very close to the digital BD. In Fig. 3 (b), the proposed hybrid schemes even outperform the digital BD. We have also observed that schemes with MMSE precoder in the baseband can achieve a better performance at low SNR, and approach the performance of schemes with ZF baseband precoder at high SNR. For sub-connected structures, a similar trend can be observed, while the performance gains of MMSE precoder over ZF precoder is more significant compared to the fully-connected case.

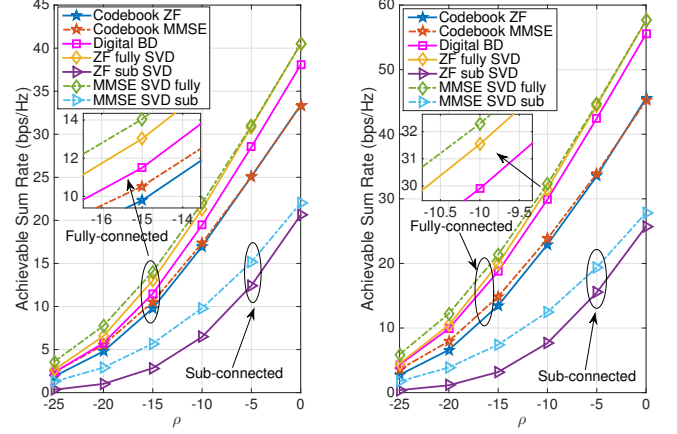


Fig. 4: Sum rate results for UPAs

In Fig. 4, the sum rate with UPAs are shown, where the quantization size for the azimuth and elevation angles is 4 bits and 3 bits respectively for the codebook-based approach. Interesting results can be observed where the proposed hybrid schemes perform even better than the digital BD scheme. It is also observed that the performance gains over the codebook based schemes become larger with the increase in the number of transmit antennas. The simulation results have shown the advantages of the proposed schemes in terms of the achievable sum rate.

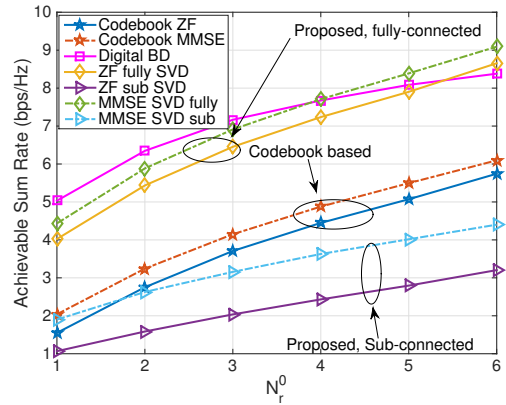


Fig. 5: Sum rates v.s. N_r^0 , $N_t=64, K=4, \text{SNR}=-10\text{dB}, L_u=6$, ULAs

Fig. 5 shows the sum capacity with respect to the increase in the number of receive antennas for ULAs. It is interesting to observe that when the number of receive antennas increases,

the performance gap between the proposed schemes and the digital BD becomes smaller. This implies that different from MU-MISO where the hybrid schemes are inferior, with multiple antennas at the receiver the hybrid schemes can outperform the conventional digital BD scheme. It is also observed that there is a constant performance gap between MMSE and ZF with the increase in the number of receive antennas.

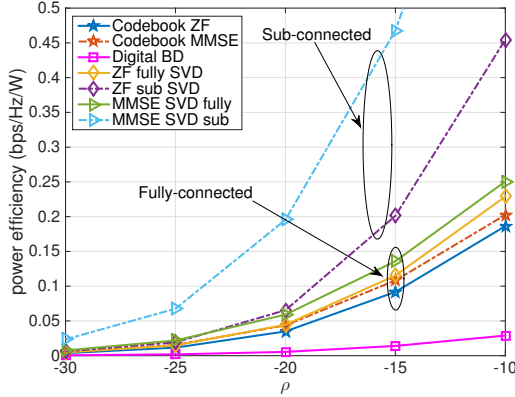


Fig. 6: Power efficiency, $N_t = 64$, $K = 4$, $N_r^0 = 4$, $L_u = 6$, UPAs

Fig. 6 compares the power efficiency between fully-connected and sub-connected structures based on (23) and (24). For digital BD, the power consumption is obtained from P_{full} in (24) by substituting N_{RF}^t with N_t . Due to the significant power consumption of RF chains, fully digital BD achieves the lowest power efficiency. The codebook-based approach in [12] is only applicable to fully-connected structures and results in the second lowest power efficiency performance due to the inferior sum rate performance. More importantly, it is clearly seen that the sub-connected structures achieve a much higher power efficiency compared to fully-connected structures, which is due to the significantly reduced number of phase shifters and the resulting reduced power consumption. It is further observed that the performance gains in the power efficiency become more dominant with the increase in the transmit SNR. Therefore, sub-connected structures are more favourable for the future green communications.

VI. CONCLUSION

In this paper, a hybrid precoding and combining design is proposed for mmWave MU-MIMO systems, where the analog domain is firstly designed, followed by the design in the baseband domain. The analog combiner is obtained based on the decomposition of the transmission channel, while we obtain the analog precoder via conjugate transposition to maximize the effective channel gains. Then, the low-dimensional baseband precoder is designed based on the resulting effective analog channel. The proposed scheme is applicable to both fully-connected structures and sub-connected structures, and the performance advantages are validated by the numerical results, where it is shown that fully-connected structures can achieve a higher sum rate, while the power efficiency is more favourable for the sub-connected structures. Future work will

focus on the designs for MU-MIMO with multiple streams per user.

ACKNOWLEDGEMENT

This work was supported by the Royal Academy of Engineering, U.K., the Engineering and Physical Sciences Research Council (EPSRC) project EP/M014150/1, and the China Scholarship Council (CSC).

REFERENCES

- [1] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?," *IEEE J. Sel. Areas in Commun.*, vol. 32, no. 6, pp. 1065-1082, June 2014.
- [2] Z. Pi, and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101-107, June 2011.
- [3] T. S. Rappaport, *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335-349, May 2013.
- [4] T. E. Bogale, and L. B. Le, "Massive MIMO and mmWave for 5G Wireless HetNet," *IEEE Veh. Tech. Mag.*, vol. 11, no. 1, pp. 64-75, Mar. 2016.
- [5] R. W. Heath Jr., N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE J. Sel. Topics in Sig. Process.*, vol. 10, no. 3, pp. 1-18, 2016.
- [6] F. Rusek, *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Sig. Process. Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013.
- [7] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.
- [8] V. Venkateswaran, and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Sig. Process.*, vol. 58, no. 8, pp. 4131-4143, April 2010.
- [9] F. Gholam, J. Via, and I. Santamaria, "Beamforming design for simplified analog antenna combining architectures," *IEEE Trans. Veh. Tech.*, vol. 60, no. 5, pp. 2373-2378, April 2011.
- [10] O. El 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.
- [11] C. Chen, "An Iterative Hybrid Transceiver Design Algorithm for Millimeter Wave MIMO Systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 285-288, June 2015.
- [12] A. Alkhateeb, G. Leus, and R. Heath Jr., "Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481-6494, July 2015.
- [13] D. H. N. Nguyen, L. B. Le, and T. Le-Ngoc, "Hybrid MMSE Precoding for mmWave Multiuser MIMO Systems," in *Proc. IEEE ICC 2016*, Kuala Lumpur, May 2016.
- [14] Z. Li, S. Han, and A. F. Molisch, "Hybrid Beamforming Design for Millimeter-Wave Multi-User Massive MIMO Downlink," in *Proc. IEEE ICC 2016*, Kuala Lumpur, May 2016.
- [15] V. Venkateswaran, F. Pivitt, and L. Guan, "Hybrid RF and digital beamformer for cellular networks: Algorithms, microwave architectures and measurements," *IEEE Trans. on Microwave Theory and Techniques*, accepted for publication, 2016. [Online]. Available: <http://arxiv.org/abs/1510.02822>.
- [16] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating Minimization Algorithms for Hybrid Precoding in Millimeter Wave MIMO Systems," *IEEE J. Sel. Topics in Sig. Process.*, vol. 10, no. 3, pp. 485-500, April 2016.
- [17] X. Gao, L. Dai, S. Han, C. I, and R. W. Heath Jr., "Energy-Efficient Hybrid Analog and Digital Precoding for MmWave MIMO Systems with Large Antenna Arrays," *IEEE J. Sel. Areas in Commun.*, vol. 34, no. 4, pp. 998-1009, Mar. 2016.
- [18] G. H. Golub, and C. F. Van Loan, *Matrix computations*. JHU Press, 2012.
- [19] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath Jr., "Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches?" *IEEE Access*, vol. 4, pp. 247-267, Jan. 2016.
- [20] Q. H. Spencer, A. L. Swindlehurst, and M. Harardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Sig. Process.*, vol. 52, no. 2, pp. 461-471, Feb. 2004.