

Received February 5, 2017, accepted March 25, 2017, date of publication March 28, 2017, date of current version June 7, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2688345

Hybrid Precoding-Beamforming Design With Hadamard RF Codebook for mmWave Large-Scale MIMO Systems

**HUSSEIN SELEEM, (Student Member, IEEE),
AHMED IYANDA SULYMAN, (Senior Member, IEEE),
AND ABDULHAMEED ALSANIE**

Electrical Engineering Department, King Saud University, Riyadh 11421, Saudi Arabia

Corresponding author: Hussein Seleem (hseleem@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research, College of Engineering Center, King Saud University.

ABSTRACT This paper proposes a hybrid structure for multi-stream large-scale multi-input multi-output (MIMO) beamforming systems, in single-user scenario, using a Hadamard radio frequency (RF) codebook with low-bit resolution phase shifters. We show that Hadamard transform can be used in RF beamsteering/beamcombining to achieve better performance in terms of average achievable spectral efficiency and low hardware cost using 1- or 2-b resolution APSs. In contrast, the state-of-the-art RF codebook designs available in the literature requires more than 7-b resolution to achieve the same performance as the proposed scheme, for large antenna arrays with up to 256 elements. The performance gains of the proposed RF codebook design is thoroughly investigated using MATLAB simulations for typical mmWave MIMO system, and the simulation results are closely verified by the analytical expressions.

INDEX TERMS Millimeter waves, large-scale antenna arrays, hybrid precoding, finite resolution analog phase shifters, RF codebook design, compressed sensing, orthogonal matching pursuit, sparse reconstruction.

I. INTRODUCTION

Millimeter wave (mmWave) frequencies (30-300 GHz) can facilitate high data rates over cellular networks. However, mmWave channels experience order of magnitude more path loss. Therefore, multi-input multi-output (MIMO) and beamforming systems are indispensable element for communications over mmWave channels due to their high achievable gains. Traditional MIMO systems employ either fully analogue beamforming at the radio frequency (RF) or fully digital precoding at the baseband where one RF chain (including up converters/down converters, digital-to-analog (DAC)/analog-to-digital (ADC) converters, mixers, and power amplifiers) is needed for each transmitting and receiving antenna. This results in prohibitively high signaling, hardware cost, and increased power consumptions when implementing large-scale MIMO systems [1].

Asymptotic analysis based on random matrix theory demonstrates that simple linear precoding structures, such as a zero forcing (ZF), are virtually optimal and comparable to nonlinear schemes in massive MIMO systems [1]. Methods to reduce the hardware costs have been the subject

of many research efforts in the literature [2]–[7]. For effective practical implementation of massive MIMO systems, the designer is thus constrained to consider hybrid precoding-beamforming strategies, which significantly reduces hardware cost by reducing the required number of RF chains from the total number of antenna elements (needed in the fully digital scheme) to at least the required number of data streams. Hybrid processing is performed through low dimensional digital baseband precoding in conjunction with analogue RF beam-steering at the transmitter and analogue RF beam-combining at the receiver, achieved via analog phase shifters (APSs).

The authors in [8]–[10] introduced analog beamforming systems that can be implemented using analog circuitry and analog phase shifters. The APSs impose a constant modulus constraint on the beamformer weights which results in poor performance compared to the fully digital precoding designs. In [11]–[16], analog beamforming was implemented based on the antenna subset selection approach, using a network of simple analog switches to limit the number of RF chains. On the other hand, hybrid beamforming was first introduced

in [8] for a point-to-point MIMO scenario, under the name of antenna soft selection. Each RF chain is connected through a network of APSs and RF adders to all antenna elements. Therefore, the number of required APSs is the product of the number of RF chains and the number of antenna elements. Recently in [7], we proposed a two-stage multiuser access scheme based on the hybrid structure which is able to serve a number of users simultaneously, with the number of required RF chains limited to the total number of users. After the seminal work of Marzetta in [17], hybrid beamforming has received more significant attentions. Marzetta introduced a new concept of massive MIMO, where the number of antenna elements at the base station (BS) and the user equipment (UE) reaches dozens or hundreds. Therefore, the hybrid structure is able to realize the optimal unconstrained precoding if there are at least two RF chains. The numbers of RF chains and analog phase shifters needed to realize the fully digital beamforming in wide band channels was addressed in [6].

Over the past three years (especially following the works in [2]–[4], [18]), dozens of papers have been published proposing various structures for hybrid beamforming. In [4] and using a compressed sensing, the authors propose a greedy algorithm called orthogonal matching pursuit (OMP) for the sparse scattering mmWave channel. It is shown that the achievable spectral efficiency maximization problem can be approximately solved by the minimization of the Frobenius norm of the difference between the optimal unconstrained SVD-based precoding and the overall hybrid beamformer. The OMP algorithm achieved a good performance when extremely large number of antennas is used at both ends. The aforementioned hybrid precoding-beamforming algorithm can reach to the optimal unconstrained performance with only high bit resolution for implementing the analog phase shifters (i.e. bit resolution should be greater than or equal $\log_2(N)$, where N is the number of antennas). However, the components required for realizing accurate phase shifters with high bit resolution can be expensive. Thus cost-effective low bit resolution phase shifters are required.

This paper considers the hybrid precoding-beamforming architecture for single-user multi-stream large-scale MIMO system for the fifth generation (5G) cellular system operating at mmWave frequency bands. We propose the design of hybrid precoding-beamforming scheme for maximizing the achievable spectral efficiency (SE) with efficient hardware implementations. We design a novel RF codebook for the RF beamsteering, and it achieves better performance in terms of average achievable SE and low hardware cost, using a 1- or 2-bit resolution APSs compared to the state-of-the-art RF codebook designs that required at least 8-bit resolution for a 256 antenna array. Moreover, the performance gain of the proposed RF codebook design is thoroughly investigate via numerical simulation.

The rest of the paper is organized as follows. Section II describes the proposed system model and formulates the problem to be discussed, while section III presents the proposed scheme for our RF codebook design and the solution

for the optimization problem. Numerical results and discussions are presented in section IV, while conclusions are presented in section V.

Notations used in the paper are as follows: upper/lower-case boldface letters denote matrices/column vectors whereas lower-case letters denote scalar; $\mathbf{A}^{[i,j]}$, \mathbf{A}^T , \mathbf{A}^* , and \mathbf{A}^H denote the $(i, j)^{th}$ element of \mathbf{A} , transpose, conjugate and conjugate transpose respectively; $\|\mathbf{A}\|_F$, $\text{tr}(\mathbf{A})$, $|\mathbf{A}|$, and $\text{diag}(\mathbf{A})$ are the Frobenius norm, trace, determinant, and the vector formed by diagonal elements of \mathbf{A} respectively; \mathbf{I}_N and $\mathbf{0}_N$ denotes the N square identity and all-zeros matrix respectively; $\mathcal{CN}(\mathbf{a}; \mathbf{A})$ denotes a complex Gaussian random vector with mean \mathbf{a} and the covariance matrix \mathbf{A} ; $\mathbb{E}[\cdot]$ denotes an expectation operator; $\mathbb{C}^{M \times N}$ and $\mathbb{R}^{M \times N}$ denotes spaces of $M \times N$ matrices with complex and real entries, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. HYBRID PRECODING-BEAMFORMING LARGE-SCALE MIMO SYSTEM MODEL

We assume the point-to-point multi-stream mmWave large-scale MIMO scenario illustrated in Fig. 1. The transmitter is equipped with N_t antennas and N_t^{RF} RF chains whereas, the receiver has N_r antennas and N_r^{RF} RF chains. The transmitter and receiver communicate via N_s data streams and to ensure effectiveness of the communications driven by the limited number of RF chains, the number of data streams is constrained to be bounded at the transmitter and receiver by $N_s \leq N_t^{RF} \leq N_t$ and $N_s \leq N_r^{RF} \leq N_r$, respectively. Without loss of generality, we assume that the number of RF chains at the transmitter and receiver are equal, $N_t^{RF} = N_r^{RF} = N^{RF}$. At the transmitter side (TX), a baseband precoding $\mathbf{F}_{BB} \in \mathbb{C}^{N^{RF} \times N_s}$ is applied in the digital domain followed by an RF beamsteering $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times N^{RF}}$ in the analog domain using analog circuitry. Each entry of the RF beamforming matrix \mathbf{F}_{RF} is normalized to satisfy $|\mathbf{F}_{RF}^{[i,j]}| = \frac{1}{\sqrt{N_t}}$, where $|(\ast)^{[i,j]}|$ denotes the modulus of the $\{i, j\}^{th}$ element of (\ast) . Moreover, \mathbf{F}_{BB} is normalized to satisfy the constrained $\|\mathbf{F}_{RF}\mathbf{F}_{BB}\|_F^2 = N_s$ on the total transmit power.

Assume a narrowband flat fading propagation environment as in [3], [4], [19]–[21], the processed received signal $\hat{\mathbf{s}} \in \mathbb{C}^{N_s \times 1}$ is given by:

$$\hat{\mathbf{s}} = \sqrt{\gamma} \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{H} \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{s} + \mathbf{W}_{BB}^H \mathbf{W}_{RF}^H \mathbf{n}, \quad (1)$$

where $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ is the information vector such that $\mathbb{E}[\mathbf{s}\mathbf{s}^*] = \frac{1}{N_s} \mathbf{I}_{N_s}$, $\gamma = \frac{\rho}{\sigma}$ is the average received signal-to-noise ratio (SNR), ρ is the average received signal power, $\mathbf{n} \in \mathbb{C}^{N_r \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the additive zero-mean circularly-symmetric complex Gaussian noise vector with unit variance σ^2 corrupting the received signal. $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix such that $\mathbb{E}[\|\mathbf{H}\|_F^2] = N_t N_r$. At the receiver side (RX), the $\mathbf{W}_{RF} \in \mathbb{C}^{N_r \times N^{RF}}$ is the RF combining matrix while $\mathbf{W}_{BB} \in \mathbb{C}^{N^{RF} \times N_s}$ is the baseband combining matrix. Similar to the RF precoder, the RF combining matrix is implemented using phase shifters with constant modulus constraint such that $|\mathbf{W}_{RF}^{[i,j]}| = \frac{1}{\sqrt{N_r}}$. To be able

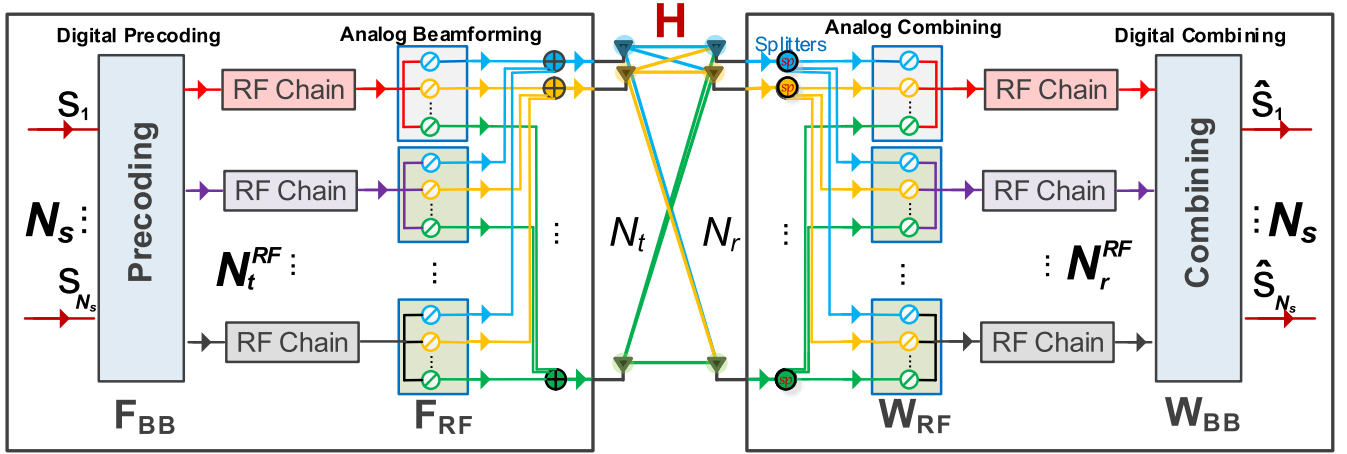


FIGURE 1. Block diagram for single-user multi-stream mmWave large-scale MIMO system with hybrid analog and digital processing. Analog beamforming and beamcombining are implemented using a simple 1- or 2-bit resolution phase shifters.

to perform the hybrid precoding-beamforming, we assume that the channel is known at both the TX and RX. The channel state information (CSI) is estimated at the RX and fed back to the TX using limited feedback [2], [4], [22], [23]. In addition, we assume perfect timing and frequency recovery. In [3], the authors developed an efficient mmWave channel estimation algorithm that leverage the geometric nature of mmWave channel. They developed a novel hierarchical multi-resolution codebook to construct training beamforming vectors with different beamwidths.

B. MILLIMETER WAVE LARGE-SCALE MIMO CHANNEL MODEL

We assume that the TX and RX both deployed square uniform planner antenna arrays (UPA). At the mmWave frequencies, the wireless channel can be characterized by a limited number of scatterers due to the high free space path loss [24], [25]. Moreover, the tightly-packed massive antenna arrays lead to high levels of correlation. The combination of high antenna correlations and sparse scattering in the mmWave channel makes many of the statistical fading distributions used in traditional MIMO analysis inaccurate for mmWave channel modeling. Therefore, based on the extended Saleh-Valenzuela model [26], we can accurately model the clustered multi-path channel encountered in the mmWave channels [4]. Using this approach the normalized physical downlink (DL) channel matrix \mathbf{H} can be written as:

$$\mathbf{H}(t) = \sqrt{\frac{N_t N_r}{N_p N_{sp}}} \sum_{i=1}^{N_p} \sum_{j=1}^{N_{sp}} \alpha_{ij} \Lambda_r(\phi_{ij}^r, \theta_{ij}^r) \Lambda_t(\phi_{ij}^t, \theta_{ij}^t) \times \mathbf{a}_r(\phi_{ij}^r, \theta_{ij}^r) \mathbf{a}_t^*(\phi_{ij}^t, \theta_{ij}^t) e^{j2\pi f_{ij}^d t} \delta(t - t_i), \quad (2)$$

where N_p is the total number of paths per channel link (i.e. the total number of scattering clusters), N_{sp} is the total number of sub-paths per path (i.e. the total number of rays per cluster), α_{ij} is the complex gain of the j^{th} sub-path in the i^{th} path, whereas $\phi_{ij}^r(\theta_{ij}^r)$ and $\phi_{ij}^t(\theta_{ij}^t)$ are its azimuth (elevation) angle

of arrival (AOA) and angle of departure (AOD) respectively. The functions $\Lambda_r(\phi_{ij}^r, \theta_{ij}^r)$ and $\Lambda_t(\phi_{ij}^t, \theta_{ij}^t)$ represent the receive and transmit antenna element gain at the angles of departure and arrival respectively. The vectors $\mathbf{a}_r(\phi_{ij}^r, \theta_{ij}^r) \in \mathbb{C}^{N_r \times 1}$ and $\mathbf{a}_t(\phi_{ij}^t, \theta_{ij}^t) \in \mathbb{C}^{N_t \times 1}$ represent the normalized receive and transmit array response vectors at an azimuth (elevation) AOA and AOD respectively. Finally, f_{ij}^d is the Doppler frequency. Since, we adopt a narrow-band clustered channel representation here, the time dependency will be omitted in our channel model.

Assume that α_{ij} are i.i.d. $\mathcal{CN}(0, \sigma_{\alpha,i}^2)$ where $\sigma_{\alpha,i}^2$ represents the average power of the i^{th} scattering cluster. The average cluster powers are such that $\sum_{i=1}^{N_p} \sigma_{\alpha,i}^2 = \sqrt{\frac{N_t N_r}{N_p N_{sp}}}$ which satisfies $\mathbb{E}[\|\mathbf{H}\|_F^2] = N_t N_r$ [27]. The azimuth (elevation) AOD and AOA within the i^{th} cluster are assumed to be randomly distributed with a uniformly-random mean cluster angle $\phi_i^t(\theta_i^t)$ and $\phi_i^r(\theta_i^r)$ respectively, and a constant angular spread (standard deviation) of $\sigma_{\phi^t}(\sigma_{\theta^t})$ and $\sigma_{\phi^r}(\sigma_{\theta^r})$ respectively. The functions $\Lambda_r(\phi_{ij}^r, \theta_{ij}^r)$ and $\Lambda_t(\phi_{ij}^t, \theta_{ij}^t)$ can be defined by the well-known far field radiation patterns for commonly used patch antennas.

Assume an N -element UPA configuration in the yz -plane, the normalized array response vectors can be written as [28]:

$$\mathbf{a}(\phi, \theta) = \frac{1}{\sqrt{N}} [1 \dots e^{jkd(m \sin(\phi) \sin(\theta) + n \cos(\theta))} \dots e^{jkd((N_y-1) \sin(\phi) \sin(\theta) + (N_z-1) \cos(\theta))}]^T, \quad (3)$$

Where $N = N_y N_z$ is the number of antenna element with N_y and N_z elements on the y and z axis respectively and $0 \leq i_y \leq N_y$ ($0 \leq i_z \leq N_z$) are the y (z) indices of an antenna elements. Whereas $k = 2\pi/\lambda$ is the wave number, $\lambda = c/f_c$ is the wavelength corresponding to the operating carrier frequency f_c . Assume the inter-element spacing $d = \lambda/2$ then $kd = \pi$. Note that $\frac{1}{\sqrt{N}}$ is the normalization factor such that $\|\mathbf{a}(\phi, \theta)\|_2^2 = 1$. In the mmWave large-scale MIMO systems, the main advantages of UPA are smaller antenna

array dimensions that leads to packing more elements in a reasonably sized array and enabling beamforming in the elevation domain (known as 3D beamforming).

The achievable spectral efficiency in (bits/s/Hz) when Gaussian symbols are transmitted over the channel and with uniform power allocations will be given by:

$$\mathcal{I}(\mathbf{s}, \hat{\mathbf{s}}) = \log_2 \left| \mathbf{I}_{N_s} + \frac{\gamma}{N_s} \mathbf{R}_n^{-1} \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \times \mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{H}^H \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}} \right| \quad (4)$$

where $\mathbf{R}_n = \mathbf{W}_{\text{BB}}^H \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}$ is the noise covariance matrix.

The design of hybrid precoders and combiners can be achieved by formulating a joint optimization problem to maximize the mutual information $\mathcal{I}(\mathbf{s}, \hat{\mathbf{s}})$ as:

$$R_{\text{HP}} = \max_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}, \mathbf{W}_{\text{RF}}, \mathbf{W}_{\text{BB}}} \mathcal{I}(\mathbf{s}, \hat{\mathbf{s}}) \quad \text{s. t. } \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = N_s \quad \mathbf{F}_{\text{RF}} \in \mathcal{F}_{\text{RF}} \quad \mathbf{W}_{\text{RF}} \in \mathcal{W}_{\text{RF}} \quad (5)$$

where $\mathcal{F}_{\text{RF}}(\mathcal{W}_{\text{RF}})$ is the set of feasible RF precoders (combiners) with all constant modulus entries. Due to the non-convexity constraints on \mathcal{F}_{RF} and \mathcal{W}_{RF} , this type of joint optimization problem is often intractable [4], [29]. Therefore, a viable approach is to construct the optimal unconstrained SVD-based precoder/combiner matrices ($\mathbf{F}_{\text{opt}}/\mathbf{W}_{\text{opt}}$) allowing a fully digital structure then try to find the practical near-optimal precoders/combiners that can be implemented in the system of Fig. 1 such that $\mathbf{F}_{\text{opt}} = \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}$ and $\mathbf{W}_{\text{opt}} = \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}$ [4]. To do this, we start by introducing the optimal unconstrained SVD-based precoding problem.

C. OPTIMAL UNCONSTRAINED SVD-BASED PRECODING DESIGN PROBLEM

The design of optimal precoder/combiner can be achieved by solving the following optimization problem [30]:

$$R_{\text{opt}} = \max_{\mathbf{F}, \mathbf{W}} \tilde{\mathcal{I}}(\mathbf{s}, \hat{\mathbf{s}}) \quad \text{s. t. } \|\mathbf{F}\|_F^2 = N_s \quad (6)$$

where R_{opt} is the optimal achievable spectral efficiency, $\tilde{\mathcal{I}}(\mathbf{s}, \hat{\mathbf{s}}) = \log_2 \left| \mathbf{I}_{N_s} + \frac{\gamma}{N_s} \mathbf{R}_n^{-1} \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{F}^H \mathbf{H}^H \mathbf{W} \right|$ and $\mathbf{R}_n = \mathbf{W}^H \mathbf{W}$. We assume that the channel matrix \mathbf{H} is well-conditioned to transmit N_s data streams such that $\text{rank}(\mathbf{H}) \geq N_s$. The optimal solution $\mathbf{F}_{\text{opt}} \in \mathbb{C}^{N_r \times N_s}$, $\mathbf{W}_{\text{opt}} \in \mathbb{C}^{N_r \times N_s}$ for the problem in (6) can be obtained by calculating the SVD of the channel matrix such that $\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$, where $\mathbf{U} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{V} \in \mathbb{C}^{N_t \times N_t}$ are unitary matrices, and $\mathbf{\Sigma} \in \mathbb{C}^{N_r \times N_t}$ is a diagonal matrix with singular values on its diagonal entry in descending order. Considering the uniform power allocation strategy, the optimal precoder/combiner are constructed with the first N_s columns of the unitary matrices

such that $\mathbf{F}_{\text{opt}} = \mathbf{V}^{\{:,1:N_s\}}$ and $\mathbf{W}_{\text{opt}} = \mathbf{U}^{\{:,1:N_s\}}$ and the optimal spectral efficiency will set the upper bound for the hybrid precoding structure in (5) and this is given by:

$$R_{\text{opt}} \stackrel{a}{=} \log_2 \left| \mathbf{I}_{N_s} + \frac{\gamma}{N_s} \mathbf{\Sigma}^{\{1:N_s, 1:N_s\}} \right| \stackrel{b}{=} \sum_{i_s=1}^{N_s} \log_2 \left(1 + \frac{\gamma}{N_s} \mathbf{\Sigma}^{\{i_s, i_s\}} \right). \quad (7)$$

D. HYBRID PRECODING-BEAMFORMING DESIGN PROBLEM

In this section we will focus on the problem formulation of the hybrid precoder \mathbf{F}_{RF} , \mathbf{F}_{BB} at the transmitter. The design can be simplified by decoupling the joint optimization problem in (4) by solving the following optimization problem at transmitter [4]:

$$\text{argmax}_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}} \tilde{\mathcal{I}}(\mathbf{s}, \mathbf{y}) \quad \text{s. t. } \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = N_s \quad \mathbf{F}_{\text{RF}} \in \mathcal{F}_{\text{RF}} \quad (8)$$

where $\mathbf{y} \in \mathbb{C}^{N_r \times 1} = \sqrt{\gamma} \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s} + \mathbf{n}$, is the received signal before processing at the receiver, and $\tilde{\mathcal{I}}(\mathbf{s}, \mathbf{y})$ is the mutual information achieved by Gaussian signaling over the mmWave channel and is given by:

$$\tilde{\mathcal{I}}(\mathbf{s}, \mathbf{y}) = \log_2 \left| \mathbf{I}_{N_r} + \frac{\gamma}{N_s} \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{H}^H \right| \quad (9)$$

The authors in [4] showed that the near optimal hybrid precoders that approximately maximize $\tilde{\mathcal{I}}(\mathbf{s}, \mathbf{y})$ in (8) can be found instead, by minimizing $\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F$. Therefore, the precoder design problem can be rewritten as:

$$\text{argmin}_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}} \|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F \quad \text{s. t. } \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = N_s \quad \mathbf{F}_{\text{RF}} \in \mathcal{F}_{\text{RF}} \quad (10)$$

This problem can be viewed as finding the projection of \mathbf{F}_{opt} onto the set of hybrid precoders of the form $\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}$ with $\mathbf{F}_{\text{RF}} \in \mathcal{F}_{\text{RF}}$. Unfortunately, the non-convexity nature of the feasible set \mathcal{F}_{RF} makes finding such a projection intractable [4].

E. HYBRID COMBINING DESIGN PROBLEM

As we assumed in section II-D that the hybrid precoder \mathbf{F}_{BB} and \mathbf{F}_{RF} is designed based on $\mathbf{F}_{\text{opt}} = \mathbf{V}^{\{:,1:N_s\}}$, the optimal combiner \mathbf{W}_{opt} could be calculated as: $\mathbf{W}_{\text{opt}} = \mathbf{U}^{\{:,1:N_s\}}$. However, the error $\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F$ can never be absolutely zero, hence the optimal combiner \mathbf{W}_{opt} deviate from $\mathbf{U}^{\{:,1:N_s\}}$. Therefore, the linear MMSE combiner will be used instead to achieve the maximum spectral efficiency with only linear combination performed. The unconstrained linear MMSE

combiner will be obtained by solving the following optimization problem:

$$\begin{aligned} & \mathbf{W}_{\text{MMSE}} \\ &= \underset{\mathbf{W}_{\text{MMSE}}}{\text{argmin}} \quad \mathbb{E}[\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2] \\ &= \frac{1}{\sqrt{\gamma}} \left(\mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{H}^H \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} + \frac{N_s}{\gamma} \mathbf{I}_{N_s} \right)^{-1} \mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H \mathbf{H}^H \quad (11) \end{aligned}$$

Once \mathbf{W}_{MMSE} is calculated, the hybrid combiners \mathbf{W}_{RF} and \mathbf{W}_{BB} can be evaluated by solving the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{W}_{\text{RF}}, \mathbf{W}_{\text{BB}}}{\text{argmin}} \quad \|\mathbf{W}_{\text{MMSE}} - \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}\|_F \\ & \text{s. t. } \mathbf{W}_{\text{RF}} \in \mathcal{W}_{\text{RF}} \quad (12) \end{aligned}$$

III. PROPOSED HYBRID PRECODING-BEAMFORMING SCHEME FOR MMWAVE CHANNEL

To provide near-optimal and low cost solutions to the problems in (10) and (12), we propose to exploit the sparse nature of the mmWave large-scale MIMO channels generated in Section II-B.

For the mmWave sparse scattering channel model with $N_t, N_r \rightarrow \infty$, the relationship between the singular and steering vectors can be expressed as: $\mathbf{v}_k = \mathbf{a}_r(\phi_k^i, \theta_k^i)$ and $\mathbf{u}_k = \mathbf{a}_r(\phi_k^r, \theta_k^r)$, $\forall k \in \{1, \dots, N_p N_{sp}\}$ [31]. Moreover, as \mathbf{V} and \mathbf{U} are unitary matrices, their column vectors can be approximated by any other unitary matrices. Among all possible unitary matrices, The Fourier and Hadamard Matrices are standard tools widely used in communications and signal processing. Both these matrices can be implemented with fast algorithms that reduce their implementation on N data points from N^2 to $N \log(N)$ operations.

A. FOURIER AND HADAMARD MATRICES

1) THE FOURIER MATRIX

$\Phi_{F_N} \in \mathbb{C}^{N \times N}$ is a unitary matrix with all its entries having constant modulus amplitude $\frac{1}{\sqrt{N}}$. We define the Fourier matrix as $\Phi_{F_N}^{(n,k)} = \frac{1}{\sqrt{N}} e^{-j \frac{2\pi}{N} nk}$, with $m \in \{0, \dots, N-1\}$ and $n \in \{0, \dots, N-1\}$. Accordingly, it can be written as:

$$\Phi_{F_N} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & \omega & \dots & \omega^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \dots & \omega^{(N-1)(N-1)} \end{bmatrix} \quad (13)$$

where $\omega = e^{-j \frac{2\pi}{N}}$.

2) THE REAL HADAMARD MATRICES

Can be considered as a complex matrices with only two phases $\{0, \pi\}$ and it's entries have only two values $\{-1, 1\}$. The real Hadamard matrix can be written in several ways, firstly as tensor Kronecker product form, secondly via a recursion and finally via their matrix elements. Assume N is a power of 2, i.e. $N = 2^K$, for some integer K .

If $\Phi_{H_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, then $\Phi_{H_N} = \Phi_{H_2} \otimes \dots \otimes \Phi_{H_2} = \otimes^K \Phi_{H_2}$. Also in terms of matrix elements, it can be written as: $\Phi_{H_N}^{(n,k)} = \frac{1}{\sqrt{N}} (-1)^{\mathbf{a}_n^T \mathbf{b}_k}$ with $m \in \{0, \dots, N-1\}$ and $n \in \{0, \dots, N-1\}$ and $\mathbf{a}_n, \mathbf{b}_k$ are vectors whose entries are the binary expansions of the matrix positions $\{n, k\}$, where $k = \sum_{l=1}^L \mathbf{a}_k(l) 2^{l-1}$ and $n = \sum_{l=1}^L \mathbf{b}_n(l) 2^{l-1}$. For instance, assuming $\{n, k\} = \{0, 3\}$ then $\mathbf{a}_0 = [0 \ 0]^T$, $\mathbf{b}_3 = [1 \ 1]^T$ such that $\mathbf{a}_0^T \mathbf{b}_3 = 0$. Therefore, $\Phi_{H_N}^{(2,3)} = \frac{1}{\sqrt{N}}(1)$. Both Fourier and Hadamard Matrices have the following properties, $\Phi_{F_N} \Phi_{F_N}^H = \mathbf{I}_N$, $\Phi_{F_N}^{-1} = \Phi_{F_N}^H$ and $\Phi_{H_N} \Phi_{H_N}^H = \mathbf{I}_N$, $\Phi_{H_N}^{-1} = \Phi_{H_N}^H$. Finally,

$$\begin{aligned} \Phi_{F_N}^{(:,k)H} \Phi_{H_N}^{(:,k')} &= \frac{1}{N} \sum_{n=0}^{N-1} e^{j \frac{2\pi}{N} nk} (-1)^{\sum_{l=1}^L \mathbf{a}_n(l) \mathbf{b}_{k'}(l)} \\ &= e^{j \left(\frac{\pi}{N} k(N-1) + \frac{\pi}{2} \sum_{l=1}^L \mathbf{b}_{k'}(l) \right)} \\ &\quad \times \prod_{l=1}^L \cos \left(\frac{\pi}{N} k 2^{l-1} + \frac{\pi}{2} \mathbf{b}_{k'}(l) \right) \quad (14) \end{aligned}$$

3) THE SEQUECNY-ORDERED COMPLEX HADAMARAD TRANSFORM (SCHT)

Is a discrete transform which is orthogonal in the complex domain and strictly confined to four complex values $\{\pm 1, \pm j\}$. Sequency is analogous to frequency in the discrete Fourier transform [32]–[35]. The SCHT matrices are generated based on the products of the row vectors of complex Rademacher matrices as follows:

$$\Phi_{H_N}^{(n,k)} = \prod_{l=0}^{K-1} \mathbf{R}_K^{(l,k) \mathbf{a}^{(l)}} \quad (15)$$

where $\mathbf{R}_K^{(l,k)} \in \mathbb{C}^{K \times N}$ is the $(l^{\text{th}}, k^{\text{th}})$ element of the complex Rademacher matrix, $\mathbf{a}^{(l)}$ is the l^{th} element of the vector whose entries are the binary expansions of the matrix position n and $K = \log_2(N)$. The complex Rademacher matrices are the discrete version of complex Rademacher functions (CRAD) and they can be generated by sampling the complex Rademacher functions as: $\mathbf{R}_K^{(l,k)} = \text{CARD}(l, \frac{4k+1}{2^{k+2}})$. The CRAD functions are defined over the time interval $0 \leq t \leq 1$ as follows:

$$\text{CRAD}(0, t) = \begin{cases} 1, & t \in [0, \frac{1}{4}) \\ j, & t \in [\frac{1}{4}, \frac{1}{2}) \\ -1, & t \in [\frac{1}{2}, \frac{3}{4}) \\ -j, & t \in [\frac{3}{4}, 1) \end{cases} \quad (16)$$

and $\text{CRAD}(0, t+1) = \text{CRAD}(0, t)$. For non-negative integer l , the CRAD functions can be defined as follows: $\text{CRAD}(l, t) = \text{CRAD}(0, 2^l t)$.

Interestingly, these Hadamard matrices can be used to approximate the array response gain of the Fourier matrices used in literature but with lower number of bit-resolution for analog phase shifters as shown in Fig. 2.

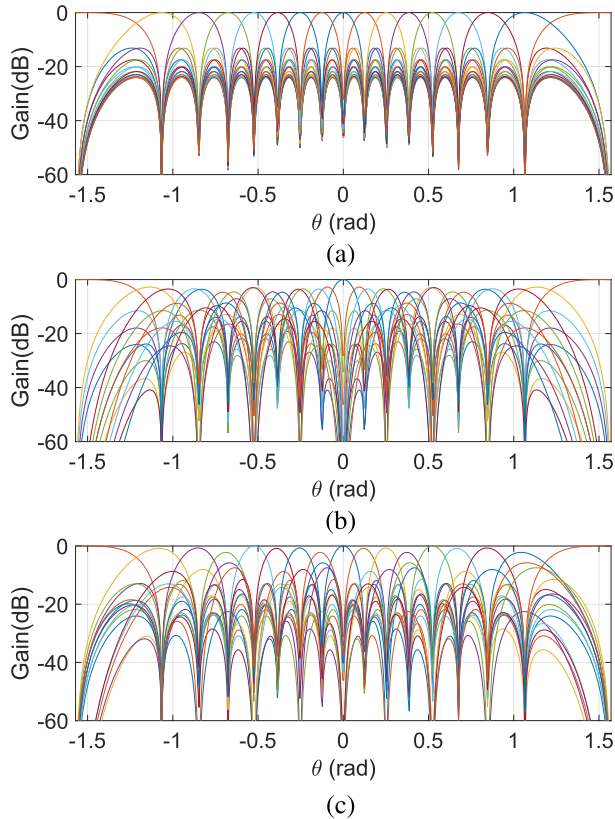


FIGURE 2. Array gains towards different spatial directions $\psi_n \in \{-\pi + \frac{\pi(n-1)}{N} \mid n = 1, \dots, N\}$ with w_{RF}^* based on the RF codebook designs with $N = 16$ uniform linear array (ULA). (a) Quantized codebook design with 4-bit resolution. (b) Proposed real Hadamard codebook design with 1-bit resolution. (c) Proposed SCHT codebook design with 2-bit resolution.

B. ACHIEVABLE SE FOR HYBRID PRECODING-BEAMFORMING STRUCTURES

The mutual information $\bar{I}(\mathbf{s}, \mathbf{y})$ in (9) achieved by the hybrid structure can be expressed as:

$$\begin{aligned} \bar{I}(\mathbf{s}, \mathbf{y}) &\stackrel{a}{=} \log_2 \left| \mathbf{I}_{N_r} + \frac{\gamma}{N_s} \mathbf{U} \Sigma \mathbf{Q} \Sigma^H \mathbf{U}^H \right| \\ &\stackrel{b}{=} \log_2 \left| \mathbf{I}_{N_r} + \frac{\gamma}{N_s} \Sigma^2 \mathbf{Q} \right| \\ &\stackrel{c}{\approx} \log_2 \left| \mathbf{I}_{N_s} + \frac{\gamma}{N_s} \Sigma^{\{1:N_s, 1:N_s\}^2} \mathbf{G}_1 \mathbf{G}_1^H \right| \\ &\stackrel{d}{\approx} \log_2 \left| \mathbf{I}_{N_s} + \frac{\gamma}{N_s} \Sigma^{\{1:N_s, 1:N_s\}^2} \right| \\ &\quad + \log_2 \left| \mathbf{I}_{N_s} - (\mathbf{I}_{N_s} - \mathbf{G}_1 \mathbf{G}_1^H) \right| \\ &\stackrel{e}{\approx} \sum_{i_s=1}^{N_s} \log_2 \left(1 + \frac{\gamma}{N_s} \Sigma^{\{i_s, i_s\}^2} \right) + \log_2 \left| \mathbf{G}_1 \mathbf{G}_1^H \right| \quad (17) \end{aligned}$$

where (a) is written in terms of SVD of \mathbf{H} , $\mathbf{Q} \in \mathbb{C}^{N_t \times N_t}$ is a positive semidefinite matrix, with $\mathbf{Q} = \mathbf{G} \mathbf{G}^H$, and $\mathbf{G} = \mathbf{V}^H \mathbf{F}_{RF} \mathbf{F}_{BB}$. Using some standard matrix identities we can rewrite (b) as shown in (c). The simplification in (c) follows from assuming that $\mathbf{G} = [\mathbf{G}_1 \ \mathbf{G}_2]^T$ where

$$\mathbf{G}_1 = \mathbf{V}^{\{:, 1:N_s\}^H} \mathbf{F}_{RF} \mathbf{F}_{BB} \text{ and } \mathbf{G}_2 = \mathbf{V}^{\{:, N_s+1:N_r\}^H} \mathbf{F}_{RF} \mathbf{F}_{BB},$$

$$\text{then } \mathbf{Q} = \begin{bmatrix} \mathbf{G}_1 \mathbf{G}_1^H & \mathbf{G}_1 \mathbf{G}_2^H \\ \mathbf{G}_2 \mathbf{G}_1^H & \mathbf{G}_2 \mathbf{G}_2^H \end{bmatrix}.$$

Assuming that $\mathbf{G}_2 \approx \mathbf{0}$, this comes from the fact that the singular values in the matrix $\Sigma^{\{N_s+1:N_r, :\}}$ are small due to the sparse nature of the mmWave channel [4].

By defining $\mathbf{A} = \frac{\gamma}{N_s} \Sigma^{\{1:N_s, 1:N_s\}^2}$ and $\mathbf{B} = \mathbf{G}_1 \mathbf{G}_1^H$ in Eq. (17c) then using the following identity: $\mathbf{I} + \mathbf{A} \mathbf{B} = (\mathbf{I} + \mathbf{A})(\mathbf{I} - (\mathbf{I} + \mathbf{A})^{-1} \mathbf{A}(\mathbf{I} - \mathbf{B}))$ and assume high SNR such that $(\mathbf{I} + \mathbf{A})^{-1} \mathbf{A} = \mathbf{I}$ we can write $\log_2 |\mathbf{I} + \mathbf{A} \mathbf{B}| = \log_2 |\mathbf{I} + \mathbf{A}| + \log_2 |(\mathbf{I} - (\mathbf{I} - \mathbf{B}))|$. Therefore, the achievable spectral efficiency can be expressed as in (17d). If we are able to design the hybrid precoder such that $\mathbf{G}_1 \approx \mathbf{I}_{N_s}$, which implies that the high-resolution RF codebook and the hybrid precoders can be made sufficiently close to the optimal unconstrained precoder, then using the following approximation: $\lim_{\mathbf{X} \rightarrow \mathbf{0}} \log_2 (\mathbf{I} - \mathbf{X}) \approx -\text{tr}(\mathbf{X})$, we obtain $\log_2 |\mathbf{I}_{N_s} - (\mathbf{I}_{N_s} - \mathbf{G}_1 \mathbf{G}_1^H)| \approx -\text{tr}(\mathbf{I}_{N_s} - \mathbf{G}_1 \mathbf{G}_1^H) = -N_s + \text{tr}(\mathbf{G}_1 \mathbf{G}_1^H) = -N_s + \|\mathbf{G}_1\|_F^2 = 0$ which indicates that the rate loss due to the hybrid precoding is equal to zero. Consequently, our problem is to maximize $\|\mathbf{G}_1\|_F^2$ or to minimize $\|\mathbf{V}^{\{:, 1:N_s\}} - \mathbf{F}_{RF} \mathbf{F}_{BB}\|_F^2$ which is equivalent to the problem in (9).

In the following sections we will develop a hybrid beamforming structures with $N_s = N^{RF}$ and calculate the lower-bound on the achievable spectral efficiency for three different cases: high-resolution PSs based on the Fourier matrix, low-resolution PSs based on the Fourier matrix, 1- and 2-bit resolution PSs based on the Hadamard matrix.

1) HIGH-RESOLUTION APSs BASED ON THE FOURIER MATRIX

Another challenge in designing the hybrid precoding structure is the finite resolution analog phase shifters. Assume q -bit resolution phase shifters are employed such that $q \geq \log_2(N_t)$. The discrete phases will be in the set $\hat{\theta}_{n,k} \in \{0, \frac{2\pi}{2^q}, \dots, \frac{2\pi}{2^q} k, \dots, \frac{2\pi}{2^q} (2^q - 1)\}$. To design the hybrid precoders $\mathbf{F}_{RF}, \mathbf{F}_{BB}$ such that it maximize $\bar{I}(\mathbf{s}, \mathbf{y})$, we seek to diagonalize $\mathbf{G}_1 = \mathbf{V}^{\{:, 1:N_s\}^H} \mathbf{F}_{RF} \mathbf{F}_{BB}$, therefore, we will start by calculating $\mathbf{V}^{\{:, 1:N_s\}^H} \mathbf{F}_{RF}$. Assume that $N_s = N^{RF}$ with large antenna arrays, and assume that $\mathbf{F}_{RF}^{\{n, k'\}} = \Phi_{F_N}^{\{n, k'\}} = e^{j\hat{\theta}_{n, k'}}$. Let $\delta_{n,k} = \hat{\theta}_{n,k} - \angle \mathbf{V}^{\{n, k\}}$ where $\frac{\pi}{2^q} \geq \delta_{n,k} \geq -\frac{\pi}{2^q}$.

$$\begin{aligned} \mathbf{v}_k^H \mathbf{f}_{RF, k'} &\stackrel{a}{=} \sum_{n=0}^{N_t} |\mathbf{V}^{\{n, k\}}| e^{-j \angle \mathbf{V}^{\{n, k\}}} e^{j \hat{\theta}_{n, k'}} \\ &\stackrel{b}{=} \sum_{n=0}^{N_t} |\mathbf{V}^{\{n, k\}}| e^{j \delta_{n, k}}, \quad k = k' \\ &\stackrel{c}{=} \sum_{n=0}^{N_t} |\mathbf{V}^{\{n, k\}}| (\cos(\delta_{n, k}) + j \sin(\delta_{n, k})) \\ &\stackrel{d}{\approx} \begin{cases} \cos(\frac{\pi}{2^q}) \sum_{n=0}^{N_t} |\mathbf{V}^{\{n, k\}}| & \text{if } k = k' \\ 0 & \text{if } k \neq k' \end{cases} \\ &\stackrel{e}{\approx} \begin{cases} \cos(\frac{\pi}{2^q}) & \text{if } k = k' \\ 0 & \text{if } k \neq k' \end{cases} \quad (18) \end{aligned}$$

where \mathbf{v}_k is the k^{th} column vector of the matrix \mathbf{V} and $\mathbf{f}_{\text{RF},k'}$ is the k'^{th} column vector of the matrix \mathbf{F}_{RF} . Equation in (c) follows from assuming high resolution phase shifters and $\theta_{n,k'}$ can be approximated as $\theta_{n,k'} = \angle \mathbf{V}^{\{n,k'\}}$. In addition, \mathbf{F}_{BB} should be diagonal matrix such that $\mathbf{F}_{\text{BB}} = \mathbf{I}_{N_s}$.

Therefore, the upper bound on the achievable rate due to high resolution phase shifters can be evaluated from (17e) and (18e) as:

$$R_{\text{F,high-res}} \approx \sum_{i_s=1}^{N_s} \log_2 \left(1 + \frac{\gamma}{N_s} \sum \{i_s, i_s\}^2 \right) + 2N_s \log_2 \left(\cos \left(\frac{\pi}{2^q} \right) \right) \quad (19)$$

where, the first term in (b) is equal to the optimal unconstrained SVD-based precoder rate.

2) LOW-RESOLUTION APSs BASED ON THE FOURIER MATRIX

In the case of q -bit resolution APSs are employed, such that $q < \log_2(N_t)$, the diagonal elements will be exist with probability $p = 2^{-(\log_2(N_t)-q)}$. Therefore,

$$\mathbf{v}_k^H \mathbf{f}_{\text{RF},k'} \approx \begin{cases} 2^{-(\log_2(N_t)-q)} \cos \left(\frac{\pi}{2^q} \right) & \text{if } k = k' \\ = 0 & \text{if } k \neq k' \end{cases} \quad (20)$$

then, the upper bound on the achievable rate due to high resolution phase shifters can be evaluated as:

$$R_{\text{F,low-res}} \approx \sum_{i_s=1}^{N_s} \log_2 \left(1 + \frac{\gamma}{N_s} \sum \{i_s, i_s\}^2 \right) + 2N_s \log_2 \left(\cos \left(\frac{\pi}{2^q} \right) \right) - 2N_s (\log_2(N_t) - q) \quad (21)$$

where, the last term in (21) is the rate loss due to the low resolution phase shifters.

3) 1- AND 2- BIT RESOLUTION APSs BASED ON THE HADAMARD MATRIX

Assume that $\mathbf{F}_{\text{RF}}^{\{n,k'\}} = \Phi_{H_N}^{\{n,k'\}}$ and $\mathbf{F}_{\text{BB}} = \mathbf{I}_{N_s}$ and assuming as mentioned before that $\mathbf{v}_k = \Phi_{F_N}^{\{:,k\}}$ we will get:

$$\mathbf{v}_k^H \mathbf{f}_{\text{RF},k'} = \sum_{n=0}^{N_t} \Phi_{F_N}^{\{n,k\}H} \Phi_{H_N}^{\{n,k'\}} \quad (22)$$

then applying (17e) where $\mathbf{G}_1 = \mathbf{V}^{\{:,1:N_s\}H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}$, we obtain an upper bound on the achievable rate using Hadamard matrix with 1-bit resolution phase shifter as:

$$\begin{aligned} R_{\text{H,1-bit}} &\stackrel{a}{\leq} \sum_{i_s=1}^{N_s} \log_2 \left(1 + \frac{\gamma}{N_s} \sum \{i_s, i_s\}^2 \right) \\ &\quad + \log_2 \left(\sum_{i_s=1}^{N_s} e^{j \left(\frac{\pi}{N} i_s (N-1) + \frac{\pi}{2} \sum_{l=1}^L \mathbf{b}_{i_s}(l) \right)} \right. \\ &\quad \quad \left. \times \prod_{l=1}^L \cos \left(\frac{\pi}{N} i_s 2^{l-1} + \frac{\pi}{2} \mathbf{b}_{i_s}(l) \right) \right) \\ &\stackrel{b}{\leq} \sum_{i_s=1}^{N_s} \log_2 \left(1 + \frac{\gamma}{N_s} \sum \{i_s, i_s\}^2 \right) \\ &\quad + 2N_s \log_2(0.45) \end{aligned} \quad (23)$$

where the value of 0.45 is the average over all the diagonal elements of \mathbf{G}_1 .

For the 2-bit Hadamard matrix case we evaluated it numerically since there is no explicit mathematical formula available for $\mathbf{v}_k^H \mathbf{f}_{\text{RF},k'}$.

$$R_{\text{H,2-bit}} \leq \sum_{i_s=1}^{N_s} \log_2 \left(1 + \frac{\gamma}{N_s} \sum \{i_s, i_s\}^2 \right) + 2N_s \log_2(0.76) \quad (24)$$

C. PROPOSED HYBRID PRECODING-BEAMFORMING STRUCTURE VIA ORTHOGONAL MATCHING PURSUIT (OMP)

By exploiting the spatial structure of the mmWave channel and formulating the hybrid precoding design problem in (10) as a sparse reconstruction problem of the optimal unconstrained SVD-based precoder, \mathbf{F}_{opt} , the hybrid precoders can be obtained by restricting \mathcal{F}_{RF} to be the set of column vectors of Hadamard matrix, Φ_{H_N} and solving:

$$\begin{aligned} &\underset{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}}{\text{argmin}} \|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F \\ &\text{s. t. } \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = N_s \\ &\quad \mathbf{f}_{\text{RF},k} \in \{\Phi_{H_N}^{\{:,k\}}, \forall k\} \end{aligned} \quad (25)$$

therefore, the problem aims to select the best N^{RF} column vectors of Φ_{H_N} that correspond to column vectors of \mathbf{F}_{opt} and then finding their optimal baseband combination that minimizes $\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|$. By embedding the constraint on $\mathbf{f}_{\text{RF},k}$ into the optimization objective and solving the equivalent problem:

$$\begin{aligned} &\underset{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}}{\text{argmin}} \|\mathbf{F}_{\text{opt}} - \Phi_{H_N} \hat{\mathbf{F}}_{\text{BB}}\|_F \\ &\text{s. t. } \|\Phi_{H_N} \hat{\mathbf{F}}_{\text{BB}}\|_F^2 = N_s \\ &\quad \|\text{diag}(\hat{\mathbf{F}}_{\text{BB}} \hat{\mathbf{F}}_{\text{BB}}^H)\|_0 = N^{RF} \end{aligned} \quad (26)$$

where $\hat{\mathbf{F}}_{\text{BB}} \in \mathbb{C}^{N_t \times N_s}$ and $\Phi_{H_N} \in \mathbb{C}^{N_t \times N_t}$ are auxiliary variables. The second constraint is the sparsity constraint which state that $\hat{\mathbf{F}}_{\text{BB}}$ cannot have more than N^{RF} nonzero rows. Therefore, only the best N^{RF} columns of the matrix Φ_{H_N} will be selected. Accordingly, the hybrid baseband and RF precoders will be given by the N^{RF} nonzero rows of $\hat{\mathbf{F}}_{\text{BB}}$ and the corresponding N^{RF} column of Φ_{H_N} , respectively. This optimization problem is equivalent to the problem of sparse signal recovery with multiple measurement vectors (MMV). Therefore, a well-known numerical algorithm based on matching pursuit [4] will be used here. The orthogonal matching pursuit (OMP) algorithm is presented in Algorithm 1. In step 4 and 5 of Algorithm 1, we find the column vectors from the dictionary matrix Φ_{H_N} , that corresponds to the maximum projection of the optimal precoding matrix, \mathbf{F}_{opt} . In step 6, we append the selected column vectors to the RF precoder, \mathbf{F}_{RF} . Then in step 7, the nonzero entries of $\hat{\mathbf{F}}_{\text{BB}}$ are obtained by solving a least square (LS) problem using pseudo-inverse of \mathbf{F}_{RF} . In step 8, the contribution of the

TABLE 1. Hardware cost at TX side for the various MIMO-beamforming structures assuming $N^{RF} = 8$ RF chains and $N_s = 4$ data streams in 256×16 mmWave large-scale MIMO system.

Beamforming Scheme	# APSs	APSs resolution	# RF Chain	Digital Complexity (flops)
Analog beamforming	$N_t \times N_s$	at least 8 bits	N_s	-
Optimal unconstrained SVD	-	-	N_t	X
Hybrid beamforming in [4]	$N_t \times N^{RF}$	at least 8 bits	N^{RF}	$X + N^{RF} N_t^2 (\frac{4}{3} N_t - 1 + N^{RF} N_t (15 N_t - 4) + N^{RF} (8 N_t - 2))$
Proposed hybrid structure	$N_t \times N^{RF}$	1- or 2-bits	N^{RF}	$X + N^{RF} N_t^2 (\frac{4}{3} N_t - 1 + N^{RF} N_t (15 N_t - 4) + N^{RF} (8 N_t - 2))$

selected vectors are removed. Finally, step 10 ensures that the transmit power constraint is exactly satisfied.

Algorithm 1 OMP Algorithm for Hybrid Design

```

1: Input:  $\mathbf{F}_{opt}, \Phi_{H_N}, N^{RF}$ 
2: Initialization:  $\mathbf{F}_{RF} = [ ]$ ,  $\mathbf{F}(1) = \mathbf{F}_{opt}$ 
3: for  $n = 1 : N^{RF}$  do
4:    $\Psi = \Phi_{H_N}^* \mathbf{F}(n)$ ,
5:    $k = \text{argmax} [\text{diag}(\Psi \Psi^*)]$ 
6:    $\mathbf{F}_{RF} = [\mathbf{F}_{RF} | \Phi_{H_N}^{(:,k)}]$ 
7:    $\mathbf{F}_{BB} = \mathbf{F}_{RF} \setminus \mathbf{F}_{opt}$ ,
8:    $\mathbf{F}(n+1) = \frac{\mathbf{F}_{opt} - \mathbf{F}_{RF} \mathbf{F}_{BB}}{\|\mathbf{F}_{opt} - \mathbf{F}_{RF} \mathbf{F}_{BB}\|_F}$ , Update the residual
9: end for
10:  $\mathbf{F}_{BB} = \sqrt{N_s} \frac{\mathbf{F}_{BB}}{\|\mathbf{F}_{RF} \mathbf{F}_{BB}\|_F}$ 
11: Output:  $\mathbf{F}_{RF}, \mathbf{F}_{BB}$ 

```

D. HARDWARE COMPLEXITY

Table 1 shows the hardware complexity comparison between various beamforming structures. Analog beamforming systems that can be implemented using analog circuitry and only APSs have lower hardware complexity in terms of number of RF chains and there is no digital baseband processing. On the other hand, it results in poor performance compared to the optimal unconstrained precoding (fully digital) and hybrid precoding designs [8]–[10]. Optimal unconstrained SVD-based precoding at the baseband, where N_t and N_r RF chains are needed at TX and RX sides respectively, results in prohibitively high signaling, hardware cost, and increased power consumptions when implementing large-scale MIMO systems. The proposed hybrid precoding-beamforming structure has same number of RF chains in [4], however the advantage in the proposed scheme is the lower number of bit resolution needed to implement APSs.

The computational load at the TX side in the digital baseband processing is primarily a function of the number of data streams N_s , the number of antennas N_t and the number of RF chains N^{RF} . The computational complexity considered here is expressed in terms of the total number of flops.¹ In real arithmetic, a multiplication followed by an addition needs 2 flops. With complex-valued quantities, a multiplication followed by an addition needs 8 flops. Thus, the complexity of a complex matrix multiplication is nearly 4 times its real

¹A flop stands for floating point operation. Operations such as addition, multiplication, subtraction, division and compare are considered as one flop.

counterpart. For a complex $m \times n$ matrix \mathbf{A} , we summarize the total FLOPs needed for the matrix operations as shown follow: (1) Multiplication of $m \times n$ and $n \times p$ complex matrices = $8mnp - 2mp$. (2) Pseudo-inversion of an $m \times n$ ($m \leq n$) complex matrix = $\frac{4}{3}m^3 + 7m^2n - m^2 - 2mn$. Therefore, the digital complexity in flops can be summarized in table 1 based on the step 4 and 7 from algorithm I. It is worth noting that the complexity of the proposed scheme is the same as complexity in [4]

IV. SIMULATION RESULTS AND DISCUSSION

In this section, several numerical simulations are carried out to test the performance of the proposed RF codebook in the hybrid precoding-beamforming architecture for single user multi-stream mmWave large-scale MIMO system and also to compare them with the existing schemes and the optimal unconstrained SVD-based precoding structure. The propagation environment between the transmitter and receiver is described in section II-B. It is modeled as a geometric channel with $N_p = 8$ clusters and $N_{sp} = 0$ rays. Further, we assume a square UPA antenna configuration. In the simulation we consider a 64×8 , 64×16 and 256×16 massive MIMO systems and it is assumed to operate at mmWave frequency band with carrier frequency of $f_c = 60$ GHz. The angle AOA/s/AODs are uniformly distributed in $[0, 2\pi]$. The system is assumed to have a bandwidth of 100 MHz, and with path loss exponent (PLE) of $n = 3.6$ for non-line-of sight (NLOS) scenarios which models typical peer-to-peer urban areas [36]. The signal to noise ratio which is defined as $\gamma = \frac{\rho}{\sigma^2}$ is controlled by varying the average received power ρ while keeping the noise power $\sigma^2 = 1$. The average achievable spectral efficiency is plotted versus received SNR over 1000 channel realizations.

In the first simulation, we consider 64×8 MIMO system with $N_s = 8$ data streams. In the proposed hybrid architecture, we assume that the number of RF chains at each end is $N_t^{RF} = N_r^{RF} = 8$ and one-bit and two-bit resolution analog phase shifters are used. Fig. 3 shows that the proposed RF codebook design with only 2-bit has better performance compared to hybrid beamforming structure with 5-bit resolution APSs in [4] by more than 10 bits/s/Hz spectral efficiency gain at SNR = 40 dB. Moreover, the theoretical upper-bound performance of the proposed structure is close to the rate of optimal unconstrained SVD-based precoding scheme. In addition, the gain loss of our proposed scheme from the sub-optimal scheme that is proposed with 6-bit resolution in [4] is around 2 bis/sec/Hz at SNR = 40 dB.

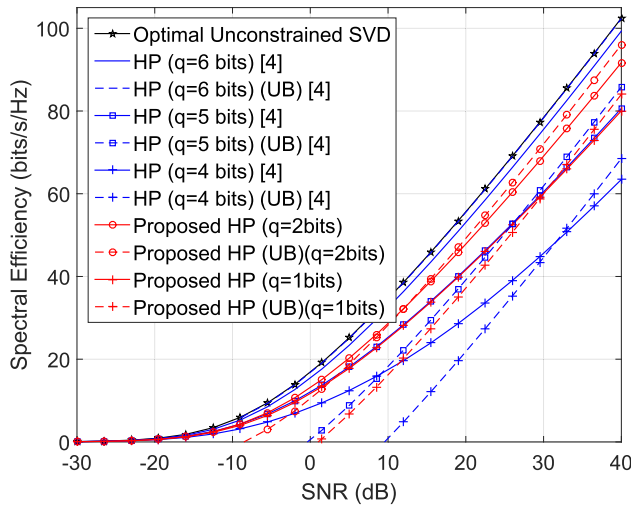


FIGURE 3. Achievable SE versus received SNR in 64×8 mmWave large-scale MIMO system with UPAs at the transmitter and receiver for various precoding-beamforming schemes. the propagation channel consists of $N_p = 8$ clusters, and $N_s = 8$ data streams are transmitted. The number of RF chains are set to $N_t^{RF} = N_r^{RF} = 8$.

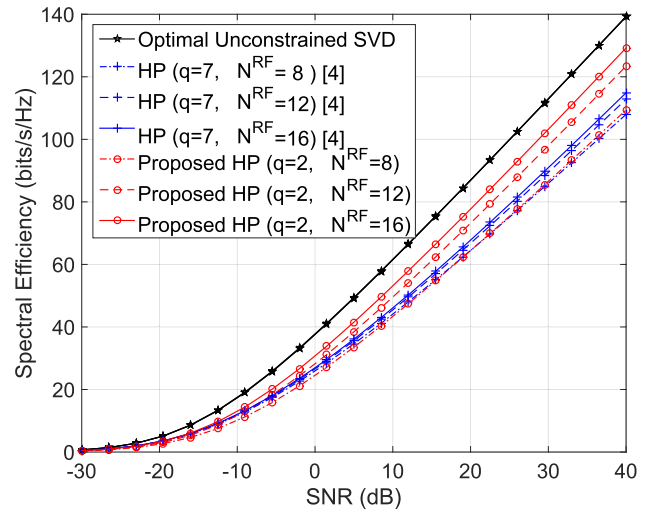


FIGURE 5. Achievable SE versus received SNR in 256×16 mmWave large-scale MIMO system with UPAs at the transmitter and receiver for three system settings at $N^{RF} = 8, 12, 16$ RF chains with the proposed 2-bit resolution RF complex Hadamard codebook. the propagation channel consists of $N_p = 8$ clusters, and $N_s = 8$ data streams are transmitted.

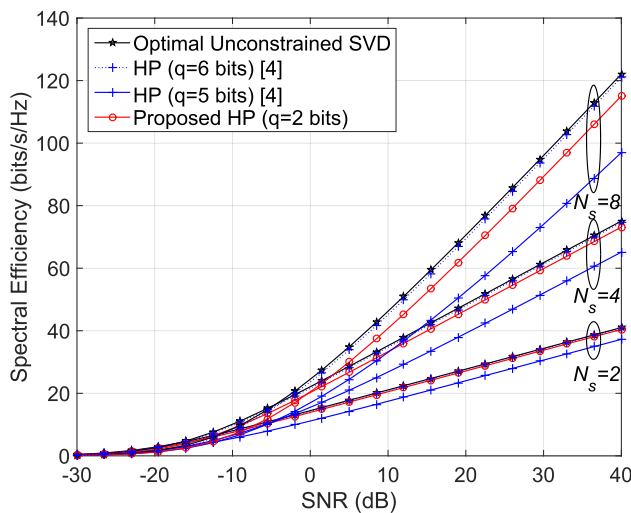


FIGURE 4. Achievable SE versus received SNR in 64×16 mmWave large-scale MIMO system with UPAs at the transmitter and receiver for various precoding-beamforming schemes. the propagation channel consists of $N_p = 8$ clusters, and $N_s = 2, 4, 8$ data streams are transmitted. The number of RF chains are set to $N_t^{RF} = N_r^{RF} = 8$.

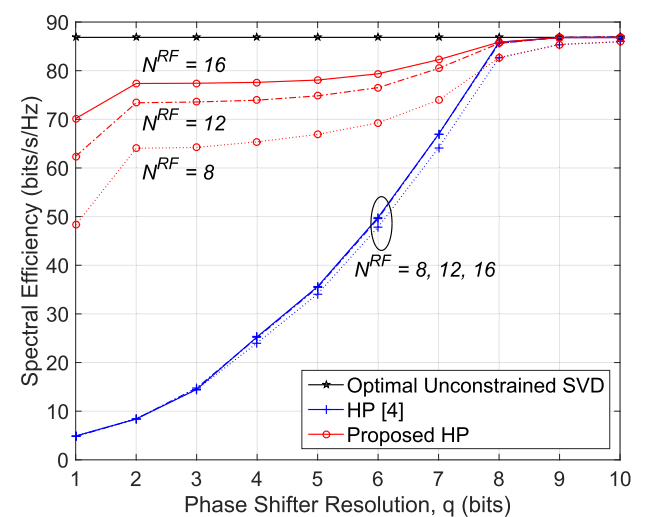


FIGURE 6. Achievable SE at $\text{SNR} = 20$ dB versus RF codebook resolution in 256×16 mmWave large-scale MIMO system with UPAs at the transmitter and receiver for various precoding-beamforming schemes. Three system settings at $N^{RF} = 8, 12, 16$ RF chains are simulated and the propagation channel consists of $N_p = 8$ clusters, and $N_s = 8$ data streams are transmitted.

Next, we analyze the performance of the proposed hybrid precoding scheme with low-bit resolution against the optimal unconstrained SVD-based structure and hybrid precoding scheme in [4] when $N_s = 2, 4, 8$ data streams are transmitted while $N_t^{RF} = N_r^{RF} = N^{RF} = 8$ RF chains over 64×16 MIMO system. As shown in Fig. 4, as we increase the number of data streams the achievable spectral efficiency is improved significantly thanks to the multiplexing gain. the gap between the proposed scheme and optimal unconstrained scheme increases as we increase the data streams and this gap can be reduced by increasing the number of RF chains as shown in Fig. 5 or increasing APSs bit resolution as

shown in Fig. 6. In addition the proposed HP scheme with two-bit resolution is again better than HP scheme with 5-bit resolution that is proposed in [4].

The proposed HP scheme with two-bit resolution is compared in Fig. 5 against the optimal scheme when $N_s = 8$, and $N^{RF} = 8, 12, 16$ RF chains at both the transmitter and receiver are employed. Although we use only one or two bit resolution analog phase shifters, the achievable spectral efficiency increases significantly with increasing the number of RF chains which To the best of the authors knowledge, this is not the case for almost all other hybrid schemes available in the literature with a finite bit resolution APSs.

Finally, Fig. 6 shows the impact of RF system limitations on the achievable spectral efficiency. Three of the proposed system models at $N_s = 8$ data streams are compared with the unconstrained system and HP system in [4], one with $N^{RF} = 8$ and the other two with $N^{RF} = 12$ and 16 RF chains, respectively. The achievable spectral efficiency is simulated with different number of quantization bits of analog phase shifters. Numerical simulation results show that the proposed hybrid precoding-beamforming scheme that include Hadamard matrices can achieve better performance with increasing the number of quantization bits and also by increasing the number of RF chains. Also these results show that only two-quantization bits may be sufficient to accomplish more than 75 % of the maximum gain.

V. CONCLUSION

In this paper, we proposed a hybrid precoding-beamforming structure for multi-stream single-user large-scale MIMO system operating at mmWave frequency bands. We show that the proposed RF Hadamard codebook with only one or two bit-resolution phase shifters can achieve performance close to the optimal unconstrained SVD-based precoding scheme with increasing the number RF chains. Furthermore, the numerical results show that the proposed architecture with only one-bit resolution analog phase shifters can improve the performance significantly better than almost all the state-of-the-art RF codebook design in the literature with high bit resolution phase shifters. The proposed structure will be very useful for 5G systems operating in mmWave bands.

REFERENCES

- [1] F. Rusek et al., "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [2] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [3] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [4] 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.
- [5] X. Zhu, Z. Wang, L. Dai, and Q. Wang, "Adaptive hybrid precoding for multiuser massive MIMO," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 776–779, Apr. 2016.
- [6] T. E. Bogale, L. B. Le, A. Haghghat, and L. Vandendorpe, "On the number of RF chains and phase shifters, and scheduling design with hybrid analog-digital beamforming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3311–3326, May 2016.
- [7] H. Seleem, A. Alsanie, and A. I. Sulyman, "Two-stage multiuser access in 5G cellular using massive MIMO and beamforming," in *Proc. Int. Conf. Cognit. Radio Oriented Wireless Netw.*, 2015, pp. 54–65.
- [8] V. Venkateswaran and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.
- [9] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.
- [10] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
- [11] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.
- [12] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection," *IEEE Commun. Mag.*, vol. 5, no. 1, pp. 46–56, Mar. 2004.
- [13] A. F. Molisch, M. Z. Win, Y.-S. Choi, and J. H. Winters, "Capacity of MIMO systems with antenna selection," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1759–1772, Jul. 2005.
- [14] M. Gharavi-Alkhansari and A. B. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 339–347, Feb. 2004.
- [15] R. W. Heath, S. Sandhu, and A. Paulraj, "Antenna selection for spatial multiplexing systems with linear receivers," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 142–144, Apr. 2001.
- [16] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive antenna selection for MIMO spatial multiplexing: Theory and algorithms," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2796–2807, Nov. 2003.
- [17] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [18] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.
- [19] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [20] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channel with partial side information," in *Proc. Conf. Rec. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, 2003, pp. 958–962.
- [21] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [22] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
- [23] J. Wang et al., "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.
- [24] H. Zhang, S. Venkateswaran, and U. Madhow, "Channel modeling and MIMO capacity for outdoor millimeter wave links," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2010, pp. 1–6.
- [25] T. S. Rappaport, F. Gutierrez, Jr., E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.
- [26] A. A. M. Saleh and R. A. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.
- [27] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, Apr. 2002.
- [28] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, NJ, USA: Wiley, 2016.
- [29] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.
- [30] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
- [31] O. El Ayach, R. W. Heath, Jr., S. Abu-Surra, S. Rajagopal, and Z. Pi, "The capacity optimality of beam steering in large millimeter wave MIMO systems," in *Proc. IEEE 13th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2012, pp. 100–104.
- [32] A. Aung, B. P. Ng, and S. Rahardja, "Sequency-ordered complex hadamard transform: Properties, computational complexity and applications," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3562–3571, Aug. 2008.

- [33] S. Kyochi and Y. Tanaka, "General factorization of conjugate-symmetric hadamard transforms," *IEEE Trans. Signal Process.*, vol. 62, no. 13, pp. 3379–3392, Jul. 2014.
- [34] J. Wu, L. Wang, G. Yang, L. Senhadji, L. Luo, and H. Shu, "Sliding conjugate symmetric sequency-ordered complex hadamard transform: Fast algorithm and applications," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 6, pp. 1321–1334, Jun. 2012.
- [35] A. Aung, B. P. Ng, and S. Rahardja, "Conjugate symmetric sequency-ordered complex hadamard transform," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2582–2593, Jul. 2009.
- [36] A. I. Sulyman, A. Alwarafy, G. R. MacCartney, T. S. Rappaport, and A. Alsanie, "Directional radio propagation path loss models for millimeter-wave wireless networks in the 28-, 60-, and 73-GHz bands," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6939–6947, Oct. 2016.



HUSSEIN SELEEM (S'08) received the B.Sc. (Hons.) and M.Sc. degrees in electronics engineering and electrical communications from Tanta University (TU), Egypt, in 2001 and 2006, respectively. He is currently pursuing the Ph.D. degree in wireless communications with the Electrical Engineering Department, King Saud University (KSU), Riyadh, Saudi Arabia. Since 2001, he has been with the Electronics Engineering and Electrical Communications Department, TU, where he is currently a Teaching Assistant. He joined the Prince Sultan Advanced Technology Research Institute, KSU, in 2010, as a Researcher, where he was involved in optical and wireless communication projects. His main research interests include signal processing for wireless and optical communications.



AHMED IYANDA SULYMAN (SM'09) received the Ph.D. degree from the Department of Electrical and Computer Engineering, Queens University, Canada, in 2006. He was a Teaching Fellow with Queens University from 2004 to 2006 and a Post-Doctoral Fellow with the Royal Military College of Canada from 2007 to 2009. He joined the Department of Electrical Engineering, King Saud University, Saudi Arabia, in 2009, where he is currently an Associate Professor. He has authored over 70 technical papers, six book chapters, and a book *Nonlinear MIMO Communication Channels* (LAP LAMBERT Academic Publishing, Saarbrücken, Deutschland/Germany: CRC Press, 2012). His research interests are broadly in wireless communications and networks, with most recent contributions in the areas of millimeter-wave 5G cellular technologies, and the Internet of Things. He has been the Session Chair and a Technical Program Committee Member in many top-tier IEEE conferences, including the most recent IEEE-ICC conference in 2016.



ABDULHAMEED ALSANIE received the B.Sc. (Hons.) and M.Sc. degrees from King Saud University in 1983 and 1987, respectively, and the Ph.D. degree from Syracuse University, NY, USA, in 1992, all in electrical engineering. He is currently an Associate Professor and the Head of the Electrical Engineering Department, King Saud University. His current research interests include wireless communications with a focus on multiple input multiple output wireless systems, space time codes, and cooperative wireless systems.

...