

# Hybrid Beamforming Based on Implicit Channel State Information for Millimeter Wave Links

Hsiao-Lan Chiang, Wolfgang Rave, Tobias Kadur, and Gerhard Fettweis, *Fellow, IEEE*

**Abstract**—Hybrid beamforming provides a promising solution to achieve high data rate transmission at millimeter waves. Implementing hybrid beamforming at a transceiver based on available channel state information is a common solution. However, many reference methods ignore the complexity of channel estimation for large antenna arrays or subsequent steps, such as the singular value decomposition of a channel matrix. To this end, we present a low-complexity scheme that exploits implicit channel knowledge to facilitate the design of hybrid beamforming for frequency-selective fading channels. The implicit channel knowledge can be interpreted as couplings between all possible pairs of analog beamforming vectors at the transmitter and receiver over the surrounding channel. Instead of calculating mutual information between large antenna arrays, we focus on small-size coupling matrices between beam patterns selected by using appropriate key parameters as performance indicators. This converts the complicated hybrid beamforming problem to a much simpler one: it amounts to collecting different sets of the large-power coupling coefficients to construct multiple alternatives for an effective channel matrix. Then, the set yielding the largest Frobenius norm (or absolute value of the determinant) of the effective channel provides the solution to the hybrid beamforming problem. It turns out that the proposed method does not require information on MIMO channel and can be simply implemented by the received correlated pilot signals that are supposed to be used for channel estimation.

**Index Terms**—millimeter wave, analog beam selection, hybrid beamforming, implicit channel state information, key parameters of hybrid beamforming gain, OFDM, MIMO.

## I. INTRODUCTION

WITH the rapid increase of data rates in wireless communications, bandwidth shortage is getting more critical. Therefore, there is a growing interest in using millimeter wave (mmWave) for future wireless communications taking advantage of the enormous amount of available spectrum at frequencies  $> 6$  GHz [1]. Measurements of mmWave channel characteristics presented in [2]-[5] show that the path loss in such an environment is very severe. In order to improve capacity and service quality, mmWave small-cell deployment together with beamforming for large antenna arrays is seen as a promising approach [6], [7]. When systems operate

Hsiao-Lan Chiang, Wolfgang Rave, Tobias Kadur, and Gerhard Fettweis are with the Vodafone Chair Mobile Communications Systems, Technische Universität Dresden, Germany, e-mail: Hsiao-lan.Chiang, Wolfgang.Rave, Tobias.Kadur, Gerhard.Fettweis@tu-dresden.de.

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 671551 (5G-XHaul) and the TUD-NEC project "mmWave Antenna Array Concept Study", a cooperation project between Technische Universität Dresden (TUD), Germany, and NEC, Japan.

This work has been presented in part at the IEEE International Communications Conference (ICC), Kansas City, MO, USA, May 2018.

at mmWave frequency bands, it is infeasible to equip each antenna with its own radio frequency (RF) chain due to high implementation cost and power consumption. Accordingly, a combination of analog beamforming (operating in passband) [8], [9] and digital beamforming (operating in baseband) [10] can be one of the low-cost solutions, and this combination is commonly called hybrid beamforming [11]-[14].

In hybrid beamforming systems, although both analog and digital beamforming matrices use the same word *beamforming*, only the former has a specific geometrical meaning in the sense of directing or collecting energy towards specific directions using the so-called array pattern of an antenna array. In contrast, the digital beamforming matrix has rather an algebraic meaning. In other words, it is more like a coefficient matrix. According to the functions of analog and digital beamforming, hybrid beamforming can be regarded as first converting an over-the-air large-scale MIMO channel matrix  $\mathbf{H}$  in the spatial domain into an effective channel  $\mathbf{H}_E$  of significantly smaller size<sup>1</sup> in the angular domain by analog beamforming vectors. Then, one can further design the weightings (i.e., the elements of the digital beamforming matrix) of the analog beamforming vectors according to some optimality criteria, for example, to maximize mutual information.

Unquestionably, it is intractable to deal with hybrid beamforming at a transmitter and a receiver simultaneously. To simplify the problem, one can assume that the channel state information (CSI) is available. Then, by utilizing the singular value decomposition (SVD) of the channel matrix, the problem of hybrid beamforming on both sides (i.e., finding the precoder and combiner) can be decoupled and formulated as two minimization problems [15]-[18]. In addition, the properties of codebooks used for the analog beamforming are incorporated into the problem as an additional constraint. For frequency-flat fading channels, a single SVD computation suffices, while more than one becomes necessary to handle multi-carrier modulation in frequency-selective fading channels. One can also decouple the transceiver by an assumption that either the transmitter or receiver employs fully digital beamforming to facilitate the problem-solving process [19]-[21]. Nevertheless, most previously proposed hybrid beamforming methods require channel knowledge and ignore the overhead of channel estimation for large-scale antenna arrays [22]-[25].

A feasible alternative that decouples the precoder and combiner is therefore introduced as follows. Generally speaking, given candidates for the analog beamforming vectors selected

<sup>1</sup>The size of  $\mathbf{H}_E$  is determined by the number of available RF chains on both sides; a more detailed description of this will be given in Section II.

from codebooks on both sides, finding the corresponding optimal digital beamforming is trivial in the sense that it is almost equivalent to the conventional fully digital beamforming apart from different power constraints [26]. As a result, the critical issue of hybrid beamforming is definitely in analog beam selection. The work in [27] explains why analog beam selection based on the power of received correlated pilot signals is equivalent to the selection method by the orthogonal matching pursuit (OMP) algorithm [28]. This happens when the analog beamforming vectors are selected from *orthogonal* codebooks (intuitively such codebooks act as complete dictionaries in terms of compressed sensing techniques). However, the performance of the analog beam selection technique based on the received power can be further improved because the corresponding effective channel  $\mathbf{H}_E$  is not necessarily well-conditioned [29], [30]. In simpler words, the factor dominating the performance of hybrid beamforming is the singular values of  $\mathbf{H}_E$  rather than the received power.

A simple cure for the problem is that one can reserve a few more candidates for the analog beamforming vectors corresponding to the large received power levels. Then, the subset of these candidates yielding the maximum throughput will provide the optimal solution to the hybrid beamforming problem. Again it is evident that the computational complexity exponentially increases as the size of the enlarged candidate set. Consequently, we have a strong motivation to find a relationship between the observations for the analog beam selection and key parameters of the hybrid beamforming gain. The relationship can be used to facilitate the process of determining the optimal analog beamforming vectors. First let us ask, what is actually the key quantity or parameter that leads to hybrid beamforming gain? Depending on the SNR, we find that it is either the Frobenius norm of the effective channel  $\mathbf{H}_E$  or the absolute value of the determinant of  $\mathbf{H}_E$ .  $\mathbf{H}_E$  can be regarded as a coupling of the channel and analog beamforming on both sides. Such coupling coefficients can be obtained by transmitting known pilot signals and used for not only the analog beam selection but also constructing alternatives for  $\mathbf{H}_E$ . Accordingly, estimates of the coupling coefficients yielding the maximum value of the key parameters give us the necessary information to optimally select the analog beamforming vectors.

The **problem statements** and **contributions** of the proposed algorithm are summarized as follows:

- 1) Most hybrid beamforming methods in the literature are implemented based on *explicit* CSI ( $\mathbf{H}$ ) but ignore the complexity of channel estimation or SVD. Therefore, this paper presents a method that uses *implicit* CSI (the received correlated pilots) to implement the hybrid beamforming and omit channel estimation.
- 2) To simplify the joint problem of the precoder and combiner, some previously proposed methods decouple these two by the assumption that either the transmitter or receiver employs fully digital beamforming. However, the assumption is not necessary. This paper shows that the precoder and combiner can be implemented simultaneously with reasonable complexity based on the estimates of the received power levels.

- 3) Compared with existing approaches, we formulate a different optimization problem by using the key parameters of the hybrid beamforming gain, which significantly alleviates the complexity of the problem.

The rest of the paper is organized as follows: Section II describes the system and mmWave frequency-selective fading channel models. Section III states the objective of hybrid beamforming problem. Based on the objective function, a hybrid beamforming algorithm based on implicit CSI is presented in Section IV. A theoretical analysis in terms of statistical properties of effective noise occurring in the proposed method is detailed in Section V. To support this analysis, simulation results are presented in Section VI, and we conclude our work in Section VII.

We use the following notations throughout this paper.

$a$	a scalar
$\mathbf{a}$	a column vector
$\mathbf{A}$	a matrix
$\mathcal{A}$	a set
$[\mathbf{A}]_{n,n}$	the $n^{\text{th}}$ diagonal element of $\mathbf{A}$
$[\mathbf{A}]_{:,1:N}$	the first $N$ column vectors of $\mathbf{A}$
$[\mathbf{A}]_{1:N,1:N}$	the $N \times N$ submatrix extracted from the upper-left corner of $\mathbf{A}$
$\mathbf{A}^*$	the complex conjugate of $\mathbf{A}$
$\mathbf{A}^H$	the Hermitian transpose of $\mathbf{A}$
$\mathbf{A}^T$	the transpose of $\mathbf{A}$
$\ \mathbf{A}\ _F$	the Frobenius norm of $\mathbf{A}$
$\det(\mathbf{A})$	the determinant of $\mathbf{A}$
$\text{vec}(\mathbf{A})$	the vectorization of $\mathbf{A}$
$[\mathbf{A} \mid \mathbf{B}]$	the horizontal concatenation
$\mathbf{A} \otimes \mathbf{B}$	the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$
$\Re(\mathbf{A}), \Im(\mathbf{A})$	the real (or imaginary) part of $\mathbf{A}$
$\mathbf{I}_N$	the $N \times N$ identity matrix
$\mathbf{0}_{N \times M}$	the $N \times M$ zero matrix
$\mathbb{E}[\cdot]$	the expectation operator

## II. SYSTEM MODEL

A system has a transmitter with a uniform linear antenna array (ULA) of  $N_T$  elements and wants to communicate  $N_S$  OFDM data streams to a receiver with an  $N_R$ -element ULA as shown in Fig. 1. At the transmitter, the  $N_T$  antenna elements connect to a precoder  $\mathbf{F}_P \mathbf{F}_B[k]$  at subcarrier  $k = 1, \dots, K$ , where  $\mathbf{F}_P \in \mathbb{C}^{N_T \times N_{RF}}$  is the analog beamforming matrix implemented in passband as part of the RF front end and  $\mathbf{F}_B[k] \in \mathbb{C}^{N_{RF} \times N_S}$  is the digital beamforming matrix in baseband. The value  $K$  specifies the number of subcarriers in one OFDM symbol, and  $N_{RF}$  denotes the number of available RF chains at both the transmitter and receiver. Costs and power consumption impose hardware constraint on the analog beamforming ( $\mathbf{F}_P$ ) so that it has fewer degrees of freedom than the digital beamforming ( $\mathbf{F}_B[k]$ ). Specifically, first,  $\mathbf{F}_P$  should be a constant matrix within (at least) one OFDM symbol, which requires certain coherence time of the channel. Second, the entries of  $\mathbf{F}_P$  have equal magnitude because analog beamformers are typically implemented by delay elements in the RF front end. The  $N_{RF}$  analog beamforming vectors of  $\mathbf{F}_P$  are selected from a predefined

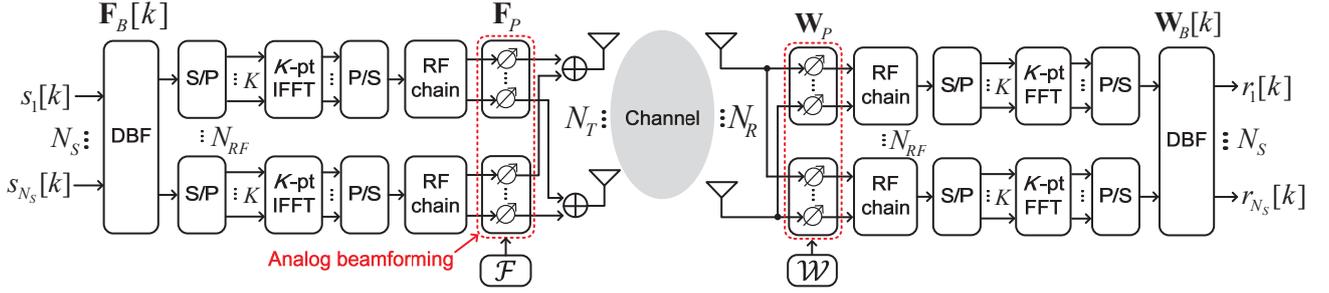


Figure 1. A MIMO-OFDM transceiver has hybrid analog and digital beamforming (DBF) structures on both sides, where each analog beamforming vector is represented by multiple phase shifters connecting to one RF chain.

codebook  $\mathcal{F} = \{\tilde{\mathbf{f}}_{n_f} \in \mathbb{C}^{N_T \times 1}, n_f = 1, \dots, N_F\}$  with the  $n_f^{\text{th}}$  member given by [8]

$$\tilde{\mathbf{f}}_{n_f} = \frac{1}{\sqrt{N_T}} \left[ 1, e^{j \frac{2\pi}{\lambda_0} \sin(\phi_{T,n_f}) \Delta_d}, \dots, e^{j \frac{2\pi}{\lambda_0} \sin(\phi_{T,n_f}) (N_T-1) \Delta_d} \right]^T, \quad (1)$$

where  $\phi_{T,n_f}$  stands for the  $n_f^{\text{th}}$  candidate for the steering angles at the transmitter,  $\Delta_d = \lambda_0/2$  is the distance between two neighboring antennas, and  $\lambda_0$  is the wavelength at the carrier frequency. At the receiver, the combiner  $\mathbf{W}_P \mathbf{W}_B[k]$  has a similar structure as the precoder, where  $\mathbf{W}_P \in \mathbb{C}^{N_R \times N_{RF}}$  and  $\mathbf{W}_B[k] \in \mathbb{C}^{N_{RF} \times N_S}$  are the analog and digital beamforming matrices respectively. Also, the columns of  $\mathbf{W}_P$  are selected from the other codebook  $\mathcal{W} = \{\tilde{\mathbf{w}}_{n_w} \in \mathbb{C}^{N_R \times 1}, n_w = 1, \dots, N_W\}$ , where the members of  $\mathcal{W}$  can be generated by the same rule as (1).

Via a coupling of the precoder, combiner, and a frequency-selective fading channel  $\mathbf{H}[k] \in \mathbb{C}^{N_R \times N_T}$ , the received signal  $\mathbf{r}[k] \in \mathbb{C}^{N_S \times 1}$  at subcarrier  $k$  can be written as

$$\begin{aligned} \mathbf{r}[k] &= \mathbf{W}_B^H[k] \mathbf{W}_P^H \mathbf{H}[k] \underbrace{\mathbf{F}_P \mathbf{F}_B[k] \mathbf{s}[k]}_{\mathbf{s}[k]} + \underbrace{\mathbf{W}_B^H[k] \mathbf{W}_P^H \mathbf{n}[k]}_{\mathbf{n}[k]} \\ &= \mathbf{W}_B^H[k] \mathbf{W}_P^H \mathbf{H}[k] \mathbf{s}[k] + \mathbf{n}[k], \end{aligned} \quad (2)$$

where  $\mathbf{s}[k] \in \mathbb{C}^{N_S \times 1}$  is the transmitted signal vector whose covariance matrix  $\mathbf{R}_s = \mathbb{E}[\mathbf{s}[k] \mathbf{s}^H[k]]$  is a diagonal matrix, and  $\mathbf{n}[k] \in \mathbb{C}^{N_R \times 1}$  is an  $N_R$ -dimensional circularly symmetric complex Gaussian random vector with mean  $\mathbf{0}_{N_R \times 1}$  and covariance matrix  $\sigma_n^2 \mathbf{I}_{N_R}$ ,  $\mathbf{n}[k] \sim \mathcal{CN}(\mathbf{0}_{N_R \times 1}, \sigma_n^2 \mathbf{I}_{N_R})$ . Furthermore, the precoded transmitted signal vector  $\mathbf{s}[k] \in \mathbb{C}^{N_T \times 1}$  and combined noise vector  $\mathbf{n}[k] \in \mathbb{C}^{N_S \times 1}$  are enforced to satisfy the following two conditions respectively: (1) constant transmit power on each subcarrier<sup>2</sup>, and (2) the entries of  $\mathbf{n}[k]$  remain i.i.d., i.e.,

$$\text{tr}(\mathbf{R}_s) = \text{tr}(\mathbf{F}_P \mathbf{F}_B[k] \mathbf{R}_s \mathbf{F}_B^H[k] \mathbf{F}_P^H) = \text{tr}(\mathbf{R}_s), \quad (3)$$

$$\mathbf{R}_n = \sigma_n^2 \mathbf{W}_B^H[k] \mathbf{W}_P^H \mathbf{W}_P \mathbf{W}_B[k] = \sigma_n^2 \mathbf{I}_{N_S}, \quad (4)$$

which can also be regarded as the power constraints on the precoder and combiner.

The properties of mmWave channels have been widely studied recently, and simulation models have been developed

<sup>2</sup>We consider a stricter condition that the constant power is allocated per subcarrier instead of per OFDM symbol for the sake of low complexity.

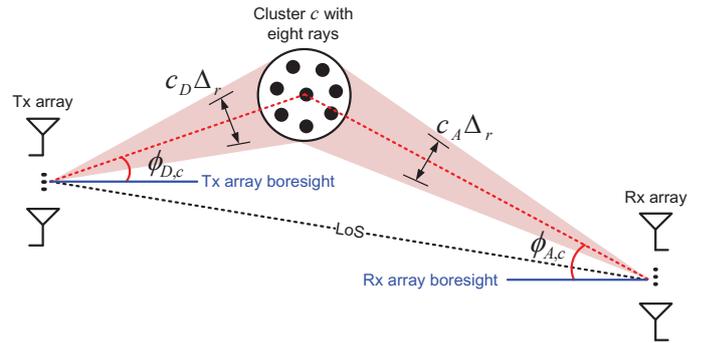


Figure 2. An example of the angular spread in a cluster characterized by its mean values ( $\phi_{D,c}$  and  $\phi_{A,c}$ ) and intra-cluster angular spreads ( $c_D \Delta_r$  and  $c_A \Delta_r$ ).

accordingly [3], [5]. The most comprehensive one can be found in [5]. Based on the references, a simplified cluster-based frequency-selective fading channel has  $C$  clusters and  $R$  rays of each cluster, where  $CR \geq N_{RF}$ . At subcarrier  $k$ , the channel matrix can be written as

$$\mathbf{H}[k] = \sqrt{\rho} \sum_{c=1}^C \sum_{r=1}^R \alpha_{c,r} \cdot e^{-j \frac{2\pi k l_{c,r}}{K}} \cdot \mathbf{a}_A(\phi_{A,c,r}) \mathbf{a}_D(\phi_{D,c,r})^H, \quad (5)$$

where the channel characteristics are given by the following parameters:

- $\rho$  stands for the average received power including the transmit power, transmit antenna gain, receive antenna gain, and path loss.
- $\alpha_{c,r} \in \mathbb{R}_{>0}$  describes the normalized inter- and intra-cluster path gain, where the mean of the difference in power between light-of-sight (LoS) and NLoS clusters is 20 dB and  $\sum_{c=1}^C \sum_{r=1}^R \alpha_{c,r}^2 = 1$ .
- Frequency-selective properties of the channel are specified in terms of normalized-quantized delays (i.e., delay indices measured in units of the sampling interval)  $l_{c,r} = \lfloor \tau_{c,r} F_S \rfloor \in \mathbb{N}_0$ , where  $\tau_{c,r}$  and  $F_S$  stand for the path delay and sampling rate respectively.
- $\phi_{D,c,r}$  is the angle of departure (AoD) of ray  $r$  in cluster  $c$ , see Fig. 2. It is characterized by the mean  $\phi_{D,c}$ , root mean square angular spread  $c_D$ , and offset angle  $\Delta_r$  for ray  $r$ , i.e.,

$$\phi_{D,c,r} = \phi_{D,c} + c_D \Delta_r, \quad (6)$$

where  $\phi_{D,c} \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$ ,  $c_D$  and  $\Delta_r$  are respectively given in [5, Table 7.5-3] and [5, Table 7.5-6]. In the same way, one can generate the angle of arrival (AoA)  $\phi_{A,c,r}$ .

- The departure array response vector  $\mathbf{a}_D(\phi_{D,c,r})$  has  $N_T$  entries of equal magnitude and is a function of  $\phi_{D,c,r}$  only. It can be written as

$$\mathbf{a}_D(\phi_{D,c,r}) = \frac{1}{\sqrt{N_T}} \begin{bmatrix} 1, e^{j\frac{2\pi}{\lambda_0} \sin(\phi_{D,c,r})\Delta_d}, \dots, \\ e^{j\frac{2\pi}{\lambda_0} \sin(\phi_{D,c,r})(N_T-1)\Delta_d} \end{bmatrix}^T. \quad (7)$$

Given an AoA, the arrival array response vector  $\mathbf{a}_A(\phi_{A,c,r})$  has a similar form as (7).

### III. PROBLEM STATEMENT

In the beamforming system, the objective of the precoder  $\mathbf{F}_P \mathbf{F}_B[k] \forall k$  and the associated combiner  $\mathbf{W}_P \mathbf{W}_B[k] \forall k$  is to achieve the maximum throughput across all subcarriers subject to the power constraints on  $\mathbf{F}_P$ ,  $\mathbf{W}_P$ ,  $\mathbf{F}_B[k]$ , and  $\mathbf{W}_B[k] \forall k$ . That is, we seek matrices that solve

$$\begin{aligned} & \max_{\mathbf{F}_P, \mathbf{W}_P, (\mathbf{F}_B[k], \mathbf{W}_B[k]) \forall k} \sum_{k=0}^{K-1} I(\mathbf{F}_P, \mathbf{W}_P, \mathbf{F}_B[k], \mathbf{W}_B[k]), \\ & \text{s.t.} \begin{cases} \mathbf{f}_{P,n_{rf}} \in \mathcal{F}, \mathbf{w}_{P,n_{rf}} \in \mathcal{W} \quad \forall n_{rf}, \\ \text{tr}(\mathbf{F}_P \mathbf{F}_B[k] \mathbf{R}_s \mathbf{F}_B^H[k] \mathbf{F}_P^H) = \text{tr}(\mathbf{R}_s) \quad \forall k, \\ \mathbf{W}_B^H[k] \mathbf{W}_P^H \mathbf{W}_P \mathbf{W}_B[k] = \mathbf{I}_{N_S} \quad \forall k, \end{cases} \end{aligned} \quad (8)$$

where  $\mathbf{f}_{P,n_{rf}}$  and  $\mathbf{w}_{P,n_{rf}}$  are respectively the  $n_{rf}^{\text{th}}$  column vectors of  $\mathbf{F}_P$  and  $\mathbf{W}_P$ , and the last two constraints are the consequences of (3) and (4). The achievable data rate of the  $k^{\text{th}}$  OFDM subchannel is given by [12], [19], [31], [32]

$$\begin{aligned} & I(\mathbf{F}_P, \mathbf{W}_P, \mathbf{F}_B[k], \mathbf{W}_B[k]) \\ & = \log_2 \det \left( \mathbf{I}_{N_S} + \mathbf{R}_n^{-1} (\mathbf{W}_B^H[k] \mathbf{W}_P^H \mathbf{H}[k] \mathbf{F}_P \mathbf{F}_B[k]) \right. \\ & \quad \left. \cdot \mathbf{R}_s (\mathbf{W}_B^H[k] \mathbf{W}_P^H \mathbf{H}[k] \mathbf{F}_P \mathbf{F}_B[k])^H \right). \end{aligned} \quad (9)$$

Moreover, we denote the solution of (8) by  $(\mathbf{F}_{P,Opt}, \mathbf{W}_{P,Opt}, (\mathbf{F}_{B,Opt}[k], \mathbf{W}_{B,Opt}[k]) \forall k)$ .

If explicit CSI is available, the problem of the precoder and combiner can be solved by exploiting the SVD of the channel matrix [15], [19], [21]. In the paper, we consider a more pragmatic approach that channel knowledge is neither given nor estimated. To efficiently get the solution of (8) without the channel knowledge, we try an alternative expression of (8), which has less probability to obtain the optimal solution, as follows: given two sets  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$  containing the candidates for  $\mathbf{F}_P$  and  $\mathbf{W}_P$ , the maximum achievable data rate of (8) is greater than or equal to

$$\max_{\substack{\mathbf{F}_P \in \mathcal{I}_{\mathcal{F}} \\ \mathbf{W}_P \in \mathcal{I}_{\mathcal{W}}}} \underbrace{\left\{ \begin{array}{l} \max_{(\mathbf{F}_B[k], \mathbf{W}_B[k]) \forall k} \sum_{k=0}^{K-1} I(\mathbf{F}_P, \mathbf{W}_P, \mathbf{F}_B[k], \mathbf{W}_B[k]) \\ \text{s.t.} \begin{cases} \text{tr}(\mathbf{F}_P \mathbf{F}_B[k] \mathbf{R}_s \mathbf{F}_B^H[k] \mathbf{F}_P^H) = \text{tr}(\mathbf{R}_s) \quad \forall k \\ \mathbf{W}_B^H[k] \mathbf{W}_P^H \mathbf{W}_P \mathbf{W}_B[k] = \mathbf{I}_{N_S} \quad \forall k \end{cases} \end{array} \right\}}_{I_{LM}(\mathbf{F}_P, \mathbf{W}_P): \text{local maximum throughput}}. \quad (10)$$

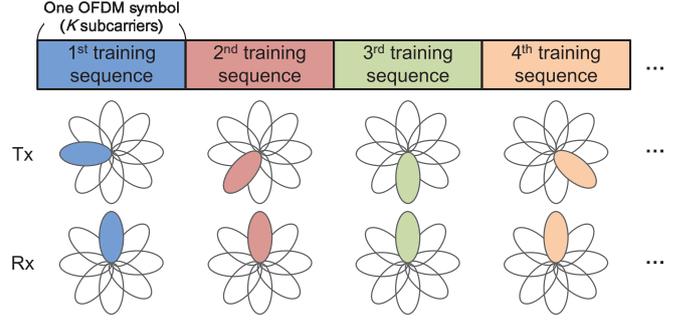


Figure 3. A training sequence of length  $K$  is used to train a beam pair.

These two versions of the hybrid beamforming problem will have the same maximum throughput if  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$  include  $\mathbf{F}_{P,Opt}$  and  $\mathbf{W}_{P,Opt}$  respectively.

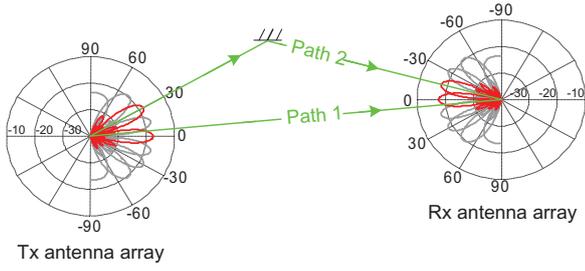
The reformulated problem in (10) becomes simpler because, given  $\mathbf{F}_P$  and  $\mathbf{W}_P$ , the *inner* problem (to obtain the local maximum throughput  $I_{LM}(\mathbf{F}_P, \mathbf{W}_P)$ ) is similar to conventional fully digital beamforming designs subject to different power constraints [10], [26]. In other words, the critical issue of the hybrid beamforming is to solve the *outer* problem by an additional maximization over all members of  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$ . Therefore, the motivation is to find  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$ , which ideally include  $\mathbf{F}_{P,Opt}$ ,  $\mathbf{W}_{P,Opt}$ , and perhaps few other candidates, and then select a pair  $(\mathbf{F}_P, \mathbf{W}_P)$  from  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$  that leads to the maximum throughput.

### IV. A HYBRID BEAMFORMING ALGORITHM BASED ON IMPLICIT CSI

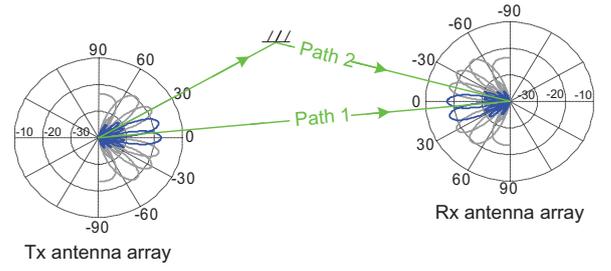
In this section, we present how to use implicit CSI to find the optimal solution to the hybrid beamforming problem. In addition, key parameters of the hybrid beamforming gain are introduced to reduce the complexity of the problem.

#### A. Initial analog beam selection

To begin with, let us see how to obtain the sets  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$  in (10) from the given codebooks  $\mathcal{F}$  and  $\mathcal{W}$ . We call this step *initial analog beam selection*. By transmitting known pilot signals, we have the received pilot signals used for the initial analog beam selection. However, these received signals include the effect of analog beamforming because the hardware-constrained analog beamforming matrices  $\mathbf{F}_P$  and  $\mathbf{W}_P$  cannot be replaced by the identity matrices. As a result, one can simply assume that all the members of the codebooks  $\mathcal{F} = \{\tilde{\mathbf{f}}_{n_f}, n_f = 1, \dots, N_F\}$  and  $\mathcal{W} = \{\tilde{\mathbf{w}}_{n_w}, n_w = 1, \dots, N_W\}$  are trained by transmitting a training sequence  $\{s[k]\}_{k=0}^{K-1}$  that satisfy  $|s[k]|^2 = 1 \forall k$ , as shown in Fig. 3. Then an observation used for the analog beam selection at subcarrier  $k$  for a specific beam pair  $(\tilde{\mathbf{f}}_{n_f}, \tilde{\mathbf{w}}_{n_w})$  can be acquired by



(a) The achievable data rate by the received power of the coupling coefficients is 2.5 bit/s/Hz.



(b) The achievable data rate by the linear combination of two analog beamforming vectors is 3 bit/s/Hz.

Figure 4. A typical example of analog beam selection by two different approaches. In the simplified two-path channel model, the AoDs are  $\{5^\circ, 30^\circ\}$ , the AoAs are  $\{5^\circ, -15^\circ\}$ , and the difference in path attenuation between path one and two amounts to 10 dB.

correlating the  $k^{\text{th}}$  received pilot with its transmitted signal

$$\begin{aligned}
 y_{n_w, n_f}[k] &= \frac{s^*[k]}{|s[k]|^2} \underbrace{\left( \tilde{\mathbf{w}}_{n_w}^H \mathbf{H}[k] \tilde{\mathbf{f}}_{n_f} s[k] + \tilde{\mathbf{w}}_{n_w}^H \mathbf{n}[k] \right)}_{\text{received pilot signal}} \\
 &= \tilde{\mathbf{w}}_{n_w}^H \mathbf{H}[k] \tilde{\mathbf{f}}_{n_f} + \underbrace{\frac{s^*[k]}{|s[k]|^2} \tilde{\mathbf{w}}_{n_w}^H \mathbf{n}[k]}_{z_{n_w, n_f}[k]} \\
 &= \tilde{\mathbf{w}}_{n_w}^H \mathbf{H}[k] \tilde{\mathbf{f}}_{n_f} + z_{n_w, n_f}[k].
 \end{aligned} \quad (11)$$

Similar observations become available on all subcarriers and the effective noise  $z_{n_w, n_f}[k] \sim \mathcal{CN}(0, \sigma_n^2)$  still has a Gaussian distribution with mean zero and variance  $\sigma_n^2$ . Also,  $z_{n_w, n_f}[k]$  is expressed as a function of  $n_w$  and  $n_f$  as the noise vector  $\mathbf{n}[k]$  is random for a trained analog beam pair  $(\tilde{\mathbf{f}}_{n_f}, \tilde{\mathbf{w}}_{n_w})$ . The observation  $y_{n_w, n_f}[k]$  can be viewed as implicit CSI, which is a coupling coefficient corresponding to a pair of analog beamforming vectors selected on both sides of the channel.

Borrowing the idea from our previous works in [27], [33], it shows that when  $\mathcal{F}$  and  $\mathcal{W}$  are orthogonal codebooks<sup>3</sup>, the sum of the power of  $K$  observations in one OFDM symbol can be directly used for the analog beam selection. Consequently,  $M$  analog beam pairs (assume that  $M \geq N_{RF}$ , which will be explained later) can be selected individually and sequentially according to the sorted received energy estimates

$$(\hat{\mathbf{f}}_m, \hat{\mathbf{w}}_m) = \arg \max_{\tilde{\mathbf{f}}_{n_f} \in \mathcal{F} \setminus \mathcal{F}', \tilde{\mathbf{w}}_{n_w} \in \mathcal{W} \setminus \mathcal{W}'} \sum_{k=0}^{K-1} |y_{n_w, n_f}[k]|^2, \quad (12)$$

where  $m = 1, \dots, M$ ,  $\mathcal{F}' = \{\tilde{\mathbf{f}}_n, n = 1, \dots, m-1\}$  and  $\mathcal{W}' = \{\tilde{\mathbf{w}}_n, n = 1, \dots, m-1\}$  are the sets consisting of the selected analog beamforming vectors from iteration 1 to  $m-1$ .

We assume that  $M \geq N_{RF}$  for the reason that the first  $N_{RF}$  selected analog beam pairs according to the sorted values of  $\sum_{k=0}^{K-1} |y_{n_w, n_f}[k]|^2$ , where  $n_w = 1, \dots, N_W$  and  $n_f = 1, \dots, N_F$ , may not be equal to the optimal solution  $(\mathbf{F}_{P, Opt}$  and  $\mathbf{W}_{P, Opt})$  because we do not yet consider the

<sup>3</sup>To be formal, an orthogonal codebook  $\mathcal{F}$  satisfies  $\frac{\langle \tilde{\mathbf{f}}_i, \tilde{\mathbf{f}}_j \rangle}{\|\tilde{\mathbf{f}}_i\|_2 \|\tilde{\mathbf{f}}_j\|_2} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$ , where  $\langle \tilde{\mathbf{f}}_i, \tilde{\mathbf{f}}_j \rangle$  denotes the inner product of the two vectors.

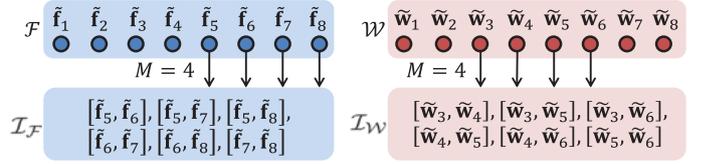


Figure 5. An example of the codebooks  $\mathcal{F}$  and  $\mathcal{W}$  and the sets  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$  consisting of  $\binom{M}{N_{RF}} = \binom{4}{2} = 6$  candidates for  $\mathbf{F}_P$  and  $\mathbf{W}_P$  respectively.

effect of digital beamforming during the analog beam selection phase. In addition, (12) is derived from the assumption that  $\mathcal{F}$  and  $\mathcal{W}$  are orthogonal codebooks. To find the optimal solution  $(\mathbf{F}_{P, Opt}$  and  $\mathbf{W}_{P, Opt})$  for any type of codebook (orthogonal or non-orthogonal), one has to further take into account linear combinations of  $N_{RF}$  analog beamforming vectors selected from  $\{\tilde{\mathbf{f}}_m \forall m\}$  and  $\{\tilde{\mathbf{w}}_m \forall m\}$  with coefficients in digital beamforming. To this end, we define two sets  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$  consisting of all combinations of  $N_{RF}$  members chosen from  $\{\tilde{\mathbf{f}}_m, m = 1, \dots, M\}$  and  $\{\tilde{\mathbf{w}}_m, m = 1, \dots, M\}$ , respectively, which can be written as

$$\begin{aligned}
 \mathcal{I}_{\mathcal{F}} &= \{\bar{\mathbf{F}}_{P, i_f}, i_f = 1, \dots, I_{\mathcal{F}}\}, \\
 \mathcal{I}_{\mathcal{W}} &= \{\bar{\mathbf{W}}_{P, i_w}, i_w = 1, \dots, I_{\mathcal{W}}\},
 \end{aligned} \quad (13)$$

where the cardinality  $I_{\mathcal{F}} = I_{\mathcal{W}} = \binom{M}{N_{RF}}$  of both sets is given by the binomial coefficient. The notations  $\bar{\mathbf{F}}_{P, i_f}$  and  $\bar{\mathbf{W}}_{P, i_w}$  respectively denote the  $i_f^{\text{th}}$  and  $i_w^{\text{th}}$  candidates for the analog beamforming matrices  $\mathbf{F}_P$  and  $\mathbf{W}_P$ . When  $M$  becomes large, there is a high probability that  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$  include the global optimum solution  $(\mathbf{F}_{P, Opt}$  and  $\mathbf{W}_{P, Opt})$ .

*Schematic example:* To illustrate the concept, let us consider a scenario with  $N_T = N_R = 8$  antenna elements, codebook sizes  $N_F = N_W = 8$ , the same orthogonal codebook  $\mathcal{F} = \mathcal{W}$  at the transmitter and receiver with the candidates for steering angles given by  $\{-90^\circ$  (or  $90^\circ$ ),  $-48.59^\circ$ ,  $-30^\circ$ ,  $-14.48^\circ$ ,  $0^\circ$ ,  $14.48^\circ$ ,  $30^\circ$ ,  $48.59^\circ\}$ , and  $N_{RF} = 2$  available RF chains to transmit  $N_S = 2$  data streams at SNR = 5 dB.

The channel realization as depicted in Fig. 4 has two paths. In Fig. 4(a), two analog beam pairs selected according to (12) steer towards these two paths (highlighted in red). Before digital beamforming comes into play, the analog beamforming vectors would be used with the same weighting. If more than

$N_{RF} = 2$  analog beam pairs are reserved, more options with digital beamforming can be explored. In this example, with  $M = 4$ , we have  $I_F = I_W = \binom{M}{N_{RF}} = \binom{4}{2} = 6$  members in both  $\mathcal{I}_F$  and  $\mathcal{I}_W$ , see Fig. 5. We enumerate them explicitly as

$$\begin{aligned}\mathcal{I}_F &= \{\bar{\mathbf{F}}_{P,i_f}, i_f = 1, \dots, 6\}, \\ \mathcal{I}_W &= \{\bar{\mathbf{W}}_{P,i_w}, i_w = 1, \dots, 6\}.\end{aligned}$$

For instance,  $\bar{\mathbf{F}}_{P,1} = [\tilde{\mathbf{f}}_5, \tilde{\mathbf{f}}_6]$  and  $\bar{\mathbf{W}}_{P,1} = [\tilde{\mathbf{w}}_3, \tilde{\mathbf{w}}_4]$ . Therefore, one can try 36 pairs,  $\{(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}) \mid \forall i_f, i_w\}$ , to determine the optimal weightings in the digital beamforming matrices and the corresponding analog beamforming matrices, which will be detailed in the following subsections. In general, there will be a competition between spatial multiplexing gain over different propagation paths and power gain available from the dominant path. In this case, the two analog beam pairs highlighted in blue in Fig. 4(b) steer to the dominant path and lead to higher spectral efficiency. However, which beamforming strategy yields higher throughput in any specific case is not clear beforehand.

### B. Digital beamforming

After the initial analog beam selection, we are in possession of the two sets  $\mathcal{I}_F$  and  $\mathcal{I}_W$  that contain the candidates for  $\mathbf{F}_P$  and  $\mathbf{W}_P$ , and the objective is to efficiently find the optimal solution. Before going into the detail of our proposed scheme, let us review the relationship between the analog and digital beamforming. Given one particular choice  $(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w})$  selected from the candidate sets  $\mathcal{I}_F$  and  $\mathcal{I}_W$ , it is clear that the goal of digital beamforming is to maximize the local maximum throughput  $I_{LM}(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w})$ , as defined in (10), with the objective function expressed as

$$\begin{aligned}& \max_{(\mathbf{F}_B[k], \mathbf{W}_B[k]) \forall k} \sum_{k=0}^{K-1} I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \mathbf{F}_B[k], \mathbf{W}_B[k]) \\ &= \sum_{k=0}^{K-1} \max_{\mathbf{F}_B[k], \mathbf{W}_B[k]} I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \mathbf{F}_B[k], \mathbf{W}_B[k]),\end{aligned}\quad (14)$$

where  $I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \mathbf{F}_B[k], \mathbf{W}_B[k])$  is given by (9) with  $\mathbf{F}_P = \bar{\mathbf{F}}_{P,i_f}$  and  $\mathbf{W}_P = \bar{\mathbf{W}}_{P,i_w}$ . As a result, the digital beamforming problem at subcarrier  $k$  can be formulated as a throughput maximization problem subject to the power constraints, which can be stated as

$$\begin{aligned}(\bar{\mathbf{F}}_{B,i}[k], \bar{\mathbf{W}}_{B,i}[k]) &= \arg \max_{\mathbf{F}_B[k], \mathbf{W}_B[k]} I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \mathbf{F}_B[k], \mathbf{W}_B[k]) \\ \text{s.t.} \quad & \begin{cases} \text{tr}(\bar{\mathbf{F}}_{P,i_f} \mathbf{F}_B[k] \mathbf{R}_s \mathbf{F}_B^H[k] \bar{\mathbf{F}}_{P,i_f}^H) = \text{tr}(\mathbf{R}_s), \\ \mathbf{W}_B^H[k] \bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w} \mathbf{W}_B[k] = \mathbf{I}_{N_S}, \end{cases}\end{aligned}\quad (15)$$

where  $i = (i_f - 1)I_W + i_w$  is an index specifying the combined members of  $\mathcal{I}_F$  and  $\mathcal{I}_W$ . To proceed, we take advantage of the mathematical results in Appendix A and have

$$\bar{\mathbf{F}}_{B,i}[k] = (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{-0.5} [\mathbf{V}_{E,i}[k]]_{:,1:N_S}, \quad (16)$$

$$\bar{\mathbf{W}}_{B,i}[k] = (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} [\mathbf{U}_{E,i}[k]]_{:,1:N_S}, \quad (17)$$

where the columns of  $\mathbf{V}_{E,i}[k]$  and  $\mathbf{U}_{E,i}[k]$  are respectively the right- and left-singular vectors of the effective channel matrix defined by

$$\mathbf{H}_{E,i}[k] \triangleq (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} \bar{\mathbf{W}}_{P,i_w}^H \mathbf{H}[k] \bar{\mathbf{F}}_{P,i_f} (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{-0.5}. \quad (18)$$

### C. Key parameters of hybrid beamforming gain

As mentioned above, given a pair of members selected from  $\mathcal{I}_F$  and  $\mathcal{I}_W$ ,  $(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w})$ , we have the corresponding optimal digital beamforming matrices  $(\bar{\mathbf{F}}_{B,i}[k], \bar{\mathbf{W}}_{B,i}[k]) \forall k$ . Accordingly, the local maximum throughput is given by

$$\begin{aligned}I_{LM}(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}) &= \sum_{k=0}^{K-1} I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \bar{\mathbf{F}}_{B,i}[k], \bar{\mathbf{W}}_{B,i}[k]) \\ &= \sum_{k=0}^{K-1} \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \frac{1}{\sigma_n^2} [\boldsymbol{\Sigma}_{E,i}^2[k]]_{n_s, n_s} [\mathbf{R}_s]_{n_s, n_s} \right),\end{aligned}\quad (19)$$

where the diagonal elements of  $\boldsymbol{\Sigma}_{E,i}[k]$  are the singular values of the effective channel  $\mathbf{H}_{E,i}[k] \stackrel{\text{SVD}}{=} \mathbf{U}_{E,i}[k] \boldsymbol{\Sigma}_{E,i}[k] \mathbf{V}_{E,i}^H[k]$ . Based on the candidate set  $\{(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}) \mid \forall i_f, i_w\}$ , the pair leading to the maximum throughput provides the best approximation of the global optimal analog beamforming matrices, that is, the solution to the hybrid beamforming problem in (10), written as

$$(\hat{\mathbf{F}}_P, \hat{\mathbf{W}}_P) = \arg \max_{\bar{\mathbf{F}}_{P,i_f} \in \mathcal{I}_F, \bar{\mathbf{W}}_{P,i_w} \in \mathcal{I}_W} I_{LM}(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}). \quad (20)$$

However, this way of solving the problem requires the SVD of  $\{\mathbf{H}_{E,i}[k]\}_{k=0}^{K-1}$  to obtain  $I_{LM}(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w})$  for each pair, which means that we have to repeat the calculation as many as  $\binom{M}{N_{RF}}^2$  times.

Alternatives that can reduce the potentially large computational burden are necessary. We ask ourselves what are the crucial parameter(s) or indicator(s) that actually determine the throughput. To answer this question, let  $\mathbf{R}_s = \frac{1}{N_S} \mathbf{I}_{N_S}$  (equal power allocation) so that the maximum achievable throughput at subcarrier  $k$  becomes

$$\begin{aligned}I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \bar{\mathbf{F}}_{B,i}[k], \bar{\mathbf{W}}_{B,i}[k]) &= \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \underbrace{\frac{1}{N_S \sigma_n^2} [\boldsymbol{\Sigma}_{E,i}^2[k]]_{n_s, n_s}}_{\triangleq \gamma} \right) \\ &= \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \gamma [\boldsymbol{\Sigma}_{E,i}^2[k]]_{n_s, n_s} \right).\end{aligned}\quad (21)$$

It is simpler to find the key parameter of the hybrid beamforming gain in the high and low SNR regimes. At low SNR

( $\gamma \rightarrow 0$ ), using the fact that  $\log(1 + \gamma x) \approx \gamma x$  as  $\gamma \rightarrow 0$ , the achievable data rate in (21) can be approximated by

$$\begin{aligned} I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \bar{\mathbf{F}}_{B,i}[k], \bar{\mathbf{W}}_{B,i}[k]) &\stackrel{\gamma \rightarrow 0}{\approx} \gamma \sum_{n_s=1}^{N_S} [\boldsymbol{\Sigma}_{E,i}^2[k]]_{n_s, n_s} \\ &\propto \sum_{n_s=1}^{N_S} [\boldsymbol{\Sigma}_{E,i}^2[k]]_{n_s, n_s} \\ &\stackrel{(a)}{\leq} \|\mathbf{H}_{E,i}[k]\|_F^2 \end{aligned} \quad (22)$$

with equality in (a) iff  $N_{RF} = N_S$ . For the case of  $N_{RF} > N_S$ ,  $\|\mathbf{H}_{E,i}[k]\|_F^2$  corresponds to the sum of all  $N_{RF}$  (instead of only the  $N_S$  strongest) eigenvalues of  $\mathbf{H}_{E,i}[k]\mathbf{H}_{E,i}^H[k]$ . Assuming that the sum of the weaker  $N_{RF} - N_S$  eigenvalues of  $\mathbf{H}_{E,i}[k]\mathbf{H}_{E,i}^H[k]$  is small, the approximation of  $\sum_{n_s=1}^{N_S} [\boldsymbol{\Sigma}_{E,i}^2[k]]_{n_s, n_s}$  by  $\|\mathbf{H}_{E,i}[k]\|_F^2$  seems to be valid for most cases of interest.

On the other hand, in the high SNR regime ( $\gamma \rightarrow \infty$ ), using  $\log(1 + \gamma x) \approx \log \gamma x$  as  $\gamma \rightarrow \infty$ , the achievable data rate in (21) is approximated by

$$\begin{aligned} I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \bar{\mathbf{F}}_{B,i}[k], \bar{\mathbf{W}}_{B,i}[k]) &\stackrel{\gamma \rightarrow \infty}{\approx} \sum_{n_s=1}^{N_S} \log_2 \left( \gamma [\boldsymbol{\Sigma}_{E,i}^2[k]]_{n_s, n_s} \right) \\ &\propto \prod_{n_s=1}^{N_S} [\boldsymbol{\Sigma}_{E,i}^2[k]]_{n_s, n_s} \\ &\stackrel{(b)}{=} |\det(\mathbf{H}_{E,i}[k])|^2, \end{aligned} \quad (23)$$

which holds with equality in (b) if  $N_{RF} = N_S$ . When the number  $CR$  of propagation paths is much larger than the number  $N_{RF}$  of RF chains ( $CR \gg N_{RF}$ ), it is reasonable to conclude that  $\det(\mathbf{H}_{E,i}[k]) \neq 0$ , i.e.,  $\text{rank}(\mathbf{H}_{E,i}[k]) = N_{RF}$ .

As we have seen, either the Frobenius norm of the effective channel matrix or absolute value of the determinant of the effective channel matrix acts as the key parameter for the system throughput. The discussion focuses on the high and low SNR regimes, and we will provide more details on approximation error in the numerical results.

#### D. Hybrid Beamforming Based on Implicit CSI

In this subsection, we will introduce how to use the coupling coefficients (or implicit CSI) to obtain the effective channel matrix  $\mathbf{H}_{E,i}[k]$ . Once we have  $\mathbf{H}_{E,i}[k]$ , the solution of the hybrid beamforming problem can be efficiently found by using the key parameters. First, let us show the effective channel presented in (18) again and approximate the elements of the matrix  $\bar{\mathbf{W}}_{P,i_w}^H \mathbf{H}[k] \bar{\mathbf{F}}_{P,i_f}$  by the coupling coefficients as the following equation

$$\begin{aligned} \mathbf{H}_{E,i}[k] &\stackrel{(18)}{=} (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} \underbrace{\bar{\mathbf{W}}_{P,i_w}^H \mathbf{H}[k] \bar{\mathbf{F}}_{P,i_f}}_{\approx \mathbf{Y}_i[k]} (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{-0.5} \\ &\approx (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} \mathbf{Y}_i[k] (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{-0.5} \\ &\triangleq \hat{\mathbf{H}}_{E,i}[k], \end{aligned} \quad (24)$$

where the elements of  $\mathbf{Y}_i[k]$  can be collected from the coupling coefficients [33]. For example, when  $\bar{\mathbf{F}}_{P,i_f} = [\tilde{\mathbf{f}}_1, \dots, \tilde{\mathbf{f}}_{N_{RF}}]$  and  $\bar{\mathbf{W}}_{P,i_w} = [\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_{N_{RF}}]$ , where  $\tilde{\mathbf{f}}_{n_{rf}}$  is the  $n_{rf}^{\text{th}}$  column of  $\mathcal{F}$  and  $\tilde{\mathbf{w}}_{n_{rf}}$  is the  $n_{rf}^{\text{th}}$  column of  $\mathcal{W}$ , one has

$$\begin{aligned} \mathbf{Y}_i[k] &= \begin{bmatrix} y_{1,1}[k] & \cdots & y_{1,N_{RF}}[k] \\ \vdots & \ddots & \vdots \\ y_{N_{RF},1}[k] & \cdots & y_{N_{RF},N_{RF}}[k] \end{bmatrix} \\ &= \bar{\mathbf{W}}_{P,i_w}^H \mathbf{H}[k] \bar{\mathbf{F}}_{P,i_f} + \mathbf{Z}[k]. \end{aligned} \quad (25)$$

Therefore, given a pair  $(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w})$  selected from  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$ , we can rapidly obtain the approximation of  $\mathbf{H}_{E,i}[k]$ , denoted by  $\hat{\mathbf{H}}_{E,i}[k]$  in (24).

In brief, the proposed solution can be stated as follows: First obtain the candidate sets ( $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$ ) and the approximation of  $\mathbf{H}_{E,i}[k]$  from the observations (or coupling coefficients)  $\{y_{n_w, n_f}[k] \forall n_w, n_f, k\}$ , and then solve the maximization problem in (20), which can be rewritten as

$$\begin{aligned} (\hat{i}_f, \hat{i}_w) &= \arg \max_{\substack{\bar{\mathbf{F}}_{P,i_f} \in \mathcal{I}_{\mathcal{F}}, \\ \bar{\mathbf{W}}_{P,i_w} \in \mathcal{I}_{\mathcal{W}}}} I_{LM}(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}) \\ &\approx \arg \max_{\substack{\bar{\mathbf{F}}_{P,i_f} \in \mathcal{I}_{\mathcal{F}}, \\ \bar{\mathbf{W}}_{P,i_w} \in \mathcal{I}_{\mathcal{W}}, \\ i = (i_f - 1)I_W + i_w}} \sum_{k=0}^{K-1} f(\hat{\mathbf{H}}_{E,i}[k]), \end{aligned} \quad (26)$$

where  $f(\hat{\mathbf{H}}_{E,i}[k])$  denotes the analog beam selection criterion using (21), (22), or (23) with the argument  $\hat{\mathbf{H}}_{E,i}[k] \stackrel{\text{SVD}}{=} \hat{\mathbf{U}}_{E,i}[k] \hat{\boldsymbol{\Sigma}}_{E,i}[k] \hat{\mathbf{V}}_{E,i}^H[k]$  given by

$$\begin{aligned} f(\hat{\mathbf{H}}_{E,i}[k]) &= \begin{cases} \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \gamma [\hat{\boldsymbol{\Sigma}}_{E,i}^2[k]]_{n_s, n_s} \right), & \text{w/o approx.} \\ \|\hat{\mathbf{H}}_{E,i}[k]\|_F^2, & \text{w/ approx. as } \gamma \rightarrow 0 \\ |\det(\hat{\mathbf{H}}_{E,i}[k])|^2, & \text{w/ approx. as } \gamma \rightarrow \infty \end{cases} \end{aligned} \quad (27)$$

Next, according to the selected index pair  $(\hat{i}_f, \hat{i}_w)$ , the selected analog and corresponding digital beamforming matrices are given by

$$\begin{aligned} \hat{\mathbf{F}}_P &= \bar{\mathbf{F}}_{P, \hat{i}_f}, \\ \hat{\mathbf{W}}_P &= \bar{\mathbf{W}}_{P, \hat{i}_w}, \\ \hat{\mathbf{F}}_B[k] &= (\hat{\mathbf{F}}_P^H \hat{\mathbf{F}}_P)^{-0.5} \left[ \hat{\mathbf{V}}_{E, \hat{i}}[k] \right]_{:, 1:N_S}, \\ \hat{\mathbf{W}}_B[k] &= (\hat{\mathbf{W}}_P^H \hat{\mathbf{W}}_P)^{-0.5} \left[ \hat{\mathbf{U}}_{E, \hat{i}}[k] \right]_{:, 1:N_S}, \end{aligned} \quad (28)$$

where  $\hat{i} = (\hat{i}_f - 1)I_W + \hat{i}_w$  and  $\hat{\mathbf{U}}_{E, \hat{i}}[k] \hat{\boldsymbol{\Sigma}}_{E, \hat{i}}[k] \hat{\mathbf{V}}_{E, \hat{i}}^H[k] = \text{SVD}(\hat{\mathbf{H}}_{E, \hat{i}}[k])$ .

The pseudocode of the proposed hybrid beamforming algorithm based on implicit CSI is shown in **Algorithm 1**. The advantages of the proposed algorithm are: (1) channel estimation for large antenna arrays can be omitted, and (2) even though the set sizes of  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$  are large, the computational overhead is minor. At low SNR, we just need to

**Algorithm 1: Hybrid beamforming based on implicit CSI****Input:**  $\{y_{n_w, n_f}[k] \forall n_w, n_f, k\}$ **Output:**  $\hat{\mathbf{F}}_P, \hat{\mathbf{W}}_P, (\hat{\mathbf{F}}_B[k], \hat{\mathbf{W}}_B[k]) \forall k$ 

1. **Part I — Initial analog beam selection**
2. Given  $\{y_{n_w, n_f}[k] \forall n_w, n_f, k\}$ , select  $M$  analog beam pairs  $(\hat{\mathbf{f}}_m, \hat{\mathbf{w}}_m)$ , where  $m = 1, \dots, M$ , by using (12).
3. Generate two candidate sets  $\mathcal{I}_{\mathcal{F}}$  and  $\mathcal{I}_{\mathcal{W}}$  based on  $\{\hat{\mathbf{f}}_m \forall m\}$  and  $\{\hat{\mathbf{w}}_m \forall m\}$ , respectively.
4. **Part II — Analog beam selection by different selection criteria**
5.  $\hat{\mathbf{H}}_{E,i}[k] = (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-0.5} \mathbf{Y}_i[k] (\overline{\mathbf{F}}_{P,i_f}^H \overline{\mathbf{F}}_{P,i_f})^{-0.5}$ , where  $\overline{\mathbf{F}}_{P,i_f} \in \mathcal{I}_{\mathcal{F}}$ ,  $\overline{\mathbf{W}}_{P,i_w} \in \mathcal{I}_{\mathcal{W}}$ , and the entries of  $\mathbf{Y}_i[k]$  are collected from  $\{y_{n_w, n_f}[k] \forall n_w, n_f\}$ .
6.  $(\hat{i}_f, \hat{i}_w) = \arg \max_{i = (i_f - 1)I_W + i_w} \sum_{k=0}^{K-1} f(\hat{\mathbf{H}}_{E,i}[k])$ , where  $f(\hat{\mathbf{H}}_{E,i}[k])$  is given by (27).
7. **Output:**  $\hat{\mathbf{F}}_P = \overline{\mathbf{F}}_{P,\hat{i}_f}$  and  $\hat{\mathbf{W}}_P = \overline{\mathbf{W}}_{P,\hat{i}_w}$ .
8. **Part III — Corresponding optimal digital beamforming**
9.  $\hat{\mathbf{U}}_{E,\hat{i}}[k] \hat{\Sigma}_{E,\hat{i}}[k] \hat{\mathbf{V}}_{E,\hat{i}}^H[k] = \text{SVD}(\hat{\mathbf{H}}_{E,\hat{i}}[k])$ , where  $\hat{i} = (\hat{i}_f - 1)I_W + \hat{i}_w$ .
10. **Output:**  $\begin{cases} \hat{\mathbf{F}}_B[k] = (\hat{\mathbf{F}}_P^H \hat{\mathbf{F}}_P)^{-0.5} [\hat{\mathbf{V}}_{E,\hat{i}}[k]]_{:,1:N_S} \\ \hat{\mathbf{W}}_B[k] = (\hat{\mathbf{W}}_P^H \hat{\mathbf{W}}_P)^{-0.5} [\hat{\mathbf{U}}_{E,\hat{i}}[k]]_{:,1:N_S} \end{cases}$

calculate the Frobenius norm of the effective channel matrices, whose elements can be easily obtained from the observations  $\{y_{n_w, n_f}[k] \forall n_w, n_f, k\}$ .

In (21), we simply assume that the transmit power is equally allocated to  $N_S$  data streams to facilitate the process of finding the best value of the key parameter. Once we find the analog and digital beamforming matrices, the global maximum throughput can be further improved by optimizing the power allocation (i.e., by a water-filling power allocation scheme [26]) for  $N_S$  data streams according to the effective channel condition.

## V. ANALYSIS OF THE PROPOSED HYBRID BEAMFORMING ALGORITHM

In the section, we focus on the statistical analysis of using the Frobenius norm of the effective channel as the key parameter at low SNR. Starting from (24),  $\hat{\mathbf{H}}_{E,i}[k]$  can be expressed as a noisy version of the true effective channel as

$$\begin{aligned} \hat{\mathbf{H}}_{E,i}[k] &\stackrel{(24)}{=} (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-0.5} \mathbf{Y}_i[k] (\overline{\mathbf{F}}_{P,i_f}^H \overline{\mathbf{F}}_{P,i_f})^{-0.5} \\ &= \underbrace{(\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-0.5} \overline{\mathbf{W}}_{P,i_w}^H \mathbf{H}[k] \overline{\mathbf{F}}_{P,i_f} (\overline{\mathbf{F}}_{P,i_f}^H \overline{\mathbf{F}}_{P,i_f})^{-0.5}}_{\mathbf{H}_{E,i}[k]} \\ &\quad + \underbrace{(\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-0.5} \mathbf{Z}[k] (\overline{\mathbf{F}}_{P,i_f}^H \overline{\mathbf{F}}_{P,i_f})^{-0.5}}_{\mathbf{Z}_{E,i}[k]} \\ &= \mathbf{H}_{E,i}[k] + \mathbf{Z}_{E,i}[k], \end{aligned} \tag{29}$$

where the multivariate distribution of the  $N_{RF}^2$ -dimensional random vector  $\text{vec}(\mathbf{Z}_{E,i}[k])$  can be written as (see Appendix B)

$$\begin{aligned} &\text{vec}(\mathbf{Z}_{E,i}[k]) \\ &\sim \mathcal{CN} \left( \mathbf{0}_{N_{RF}^2 \times 1}, \sigma_n^2 \left( (\overline{\mathbf{F}}_{P,i_f}^T \overline{\mathbf{F}}_{P,i_f}^*)^{-1} \otimes (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-1} \right) \right). \end{aligned} \tag{30}$$

From (29), we have  $\|\hat{\mathbf{H}}_{E,i}[k]\|_F^2$  given by

$$\begin{aligned} \underbrace{\|\hat{\mathbf{H}}_{E,i}[k]\|_F^2}_{\text{estimate of the key parameter}} &= \underbrace{\|\mathbf{H}_{E,i}[k]\|_F^2}_{\text{true value of the key parameter}} \\ &\quad + \underbrace{\|\mathbf{Z}_{E,i}[k]\|_F^2 + 2 \cdot \Re(\text{tr}(\mathbf{H}_{E,i}^H[k] \mathbf{Z}_{E,i}[k]))}_{\text{noise}}. \end{aligned} \tag{31}$$

To analyze the noise effect, we introduce the quantities

$$U = \|\mathbf{Z}_{E,i}[k]\|_F^2, \tag{32}$$

$$V = 2 \cdot \Re(\text{tr}(\mathbf{H}_{E,i}^H[k] \mathbf{Z}_{E,i}[k])), \tag{33}$$

and pursue the analysis of  $U$  and  $V$  for orthogonal and non-orthogonal codebooks.

### A. Orthogonal codebooks

When the columns of  $\overline{\mathbf{F}}_{P,i_f}$  and the columns of  $\overline{\mathbf{W}}_{P,i_w}$  are mutually orthogonal respectively, from (30) we know that the elements of  $\text{vec}(\mathbf{Z}_{E,i}[k])$  have the same normal distribution with mean zero and variance  $\sigma_n^2$ ,  $\text{vec}(\mathbf{Z}_{E,i}[k]) \sim \mathcal{CN}(\mathbf{0}_{N_{RF}^2 \times 1}, \sigma_n^2 \mathbf{I}_{N_{RF}^2})$ . Therefore,  $U$  is the sum of the absolute squares of  $N_{RF}^2$  i.i.d. Gaussian random variables, which follows a Gamma distribution with shape parameter  $N_{RF}^2$  and scale parameter  $\sigma_n^2$ :

$$\begin{aligned} U &= \sum_{i=1}^{N_{RF}} \sum_{j=1}^{N_{RF}} \underbrace{\Re([\mathbf{Z}_{E,i}[k]]_{i,j})^2}_{\sim \Gamma(\frac{1}{2}, \sigma_n^2)} + \underbrace{\Im([\mathbf{Z}_{E,i}[k]]_{i,j})^2}_{\sim \Gamma(\frac{1}{2}, \sigma_n^2)} \\ &\sim \Gamma(N_{RF}^2, \sigma_n^2). \end{aligned} \tag{34}$$

In addition, given one channel realization, denoted by  $\mathbf{H}'[k]$ ,  $V$  is normally distributed with mean zero and variance  $2\sigma_n^2 \|\mathbf{H}'_{E,i}[k]\|_F^2$ , where  $\mathbf{H}'_{E,i}[k] = \overline{\mathbf{W}}_{P,i_w}^H \mathbf{H}'[k] \overline{\mathbf{F}}_{P,i_f}$ .

### B. Non-orthogonal codebooks

When the columns of  $\overline{\mathbf{F}}_{P,i_f}$  or the columns of  $\overline{\mathbf{W}}_{P,i_w}$  are not mutually orthogonal, the elements of  $\text{vec}(\mathbf{Z}_{E,i}[k])$  in (30) are not i.i.d. anymore. In this case, there are no closed-form expressions for the probability distributions of  $U$  and  $V$ . Accordingly, we only derive and state  $E[U]$ ,  $\text{Var}(U)$ , and  $E[V]$  in this section. These are given by (see Appendix C)

$$E[U] = \sigma_n^2 \cdot \text{tr} \left( (\overline{\mathbf{F}}_{P,i_f}^T \overline{\mathbf{F}}_{P,i_f}^*)^{-1} \right) \text{tr} \left( (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-1} \right), \tag{35}$$

and

$$\text{Var}(U) = \text{tr}(\Psi \mathbf{R}_{z_V}) - E[U]^2, \tag{36}$$

where

$$\begin{aligned} \Psi &= \left( (\overline{\mathbf{F}}_{P,i_f}^T \overline{\mathbf{F}}_{P,i_f}^*)^{-1} \otimes (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-1} \right) \\ &\quad \otimes \left( (\overline{\mathbf{F}}_{P,i_f}^T \overline{\mathbf{F}}_{P,i_f}^*)^{-1} \otimes (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-1} \right), \end{aligned} \tag{37}$$

**Algorithm 2: Hybrid beamforming based on explicit CSI**

**Input:**  $\{\mathbf{V}[k] \forall k\}$   
**Output:**  $\tilde{\mathbf{F}}_P, \tilde{\mathbf{F}}_B[k] \forall k$

1.  $\tilde{\mathbf{F}}_P = \text{empty matrix}$
2.  $\mathbf{V}_R[k] = [\mathbf{V}[k]]_{:,1:N_S}$
3. for  $n_{rf} = 1, \dots, N_{RF}$
4.  $\tilde{\mathbf{f}}_{P,n_{rf}} = \arg \max_{\tilde{\mathbf{f}}_{n_{rf}} \in \mathcal{F}} \sum_{k=0}^{K-1} \left\| \tilde{\mathbf{f}}_{n_{rf}}^H \mathbf{V}_R[k] \right\|_F^2$
5.  $\tilde{\mathbf{F}}_P = [\tilde{\mathbf{F}}_P | \tilde{\mathbf{f}}_{P,n_{rf}}]$
6.  $\mathbf{V}_R[k] = (\mathbf{I}_{N_T} - \tilde{\mathbf{F}}_P (\tilde{\mathbf{F}}_P^H \tilde{\mathbf{F}}_P)^{-1} \tilde{\mathbf{F}}_P^H) [\mathbf{V}[k]]_{:,1:N_S}$
7.  $\mathbf{V}_R[k] = \frac{\mathbf{V}_R[k]}{\|\mathbf{V}_R[k]\|_F}$
8. end
9.  $\tilde{\mathbf{F}}_B[k] = (\tilde{\mathbf{F}}_P^H \tilde{\mathbf{F}}_P)^{-1} \tilde{\mathbf{F}}_P^H [\mathbf{V}[k]]_{:,1:N_S}$
10.  $\tilde{\mathbf{F}}_B[k] = \sqrt{N_S} \cdot \frac{\tilde{\mathbf{F}}_B[k]}{\|\tilde{\mathbf{F}}_P \tilde{\mathbf{F}}_B[k]\|_F}$

$$\mathbf{R}_{z_V} = \mathbb{E} \left[ \underbrace{(\text{vec}(\mathbf{Z}[k]) \otimes \text{vec}(\mathbf{Z}[k]))}_{\mathbf{z}_V[k] \in \mathbb{C}^{N_{RF}^4 \times 1}} \underbrace{(\text{vec}(\mathbf{Z}[k]) \otimes \text{vec}(\mathbf{Z}[k]))^H}_{\mathbf{z}_V^H[k]} \right], \quad (38)$$

and  $\mathbb{E}[V] = 0$ . Unfortunately, we did not find a closed-form expression for  $\text{Var}(V)$ . From the analysis results, it is clear that the distributions of  $U$  and  $V$  for non-orthogonal codebooks vary with the analog beamforming candidates  $\tilde{\mathbf{F}}_{P,i_f}$  and  $\tilde{\mathbf{W}}_{P,i_w}$ , which means that the estimates of  $\|\tilde{\mathbf{H}}_{E,i}[k]\|_F^2$  in (31) may become unreliable because of the non-i.i.d. noise signals.

## VI. SIMULATION RESULTS

The system parameters used in the simulations are listed below. In addition, the SNR in linear scale is given by  $\text{SNR} = \frac{\rho}{N_S \sigma_n^2}$ , where  $\rho$  is the average received power.

Number of antennas	$N_T = N_R = 32$
Number of RF chains	$N_{RF} = 2$
Number of data streams	$N_S = 2$
Length of a training sequence	$K = 512$
Number of clusters	$C = 5$ (1 LoS and 4 NLoS clusters)
Number of rays per cluster	$R = 8$

We chose the work in [15] that implements hybrid beamforming based on *explicit CSI* as a reference method for comparison and extended it from single carrier to multiple carriers. In the reference method, given the channel matrices,  $\mathbf{H}[k] \stackrel{\text{SVD}}{=} \mathbf{U}[k] \mathbf{\Sigma}[k] \mathbf{V}^H[k] \forall k$ , the goal of the precoder design is to minimize the sum of the squared Frobenius norms of the errors between the right singular vectors and the precoder across all subcarriers:

$$\begin{aligned} (\tilde{\mathbf{F}}_P, \tilde{\mathbf{F}}_B[k] \forall k) &= \arg \min_{\tilde{\mathbf{F}}_P, \tilde{\mathbf{F}}_B[k] \forall k} \sum_{k=0}^{K-1} \left\| [\mathbf{V}[k]]_{:,1:N_S} - \tilde{\mathbf{F}}_P \tilde{\mathbf{F}}_B[k] \right\|_F^2, \\ \text{s.t. } &\begin{cases} \tilde{\mathbf{f}}_{P,n_{rf}} \in \mathcal{F} \forall n_{rf}, \\ \|\tilde{\mathbf{F}}_P \tilde{\mathbf{F}}_B[k]\|_F^2 = N_S \forall k. \end{cases} \end{aligned} \quad (39)$$

The problem can be solved by the OMP [28] and the pseudocode is given in **Algorithm 2**. Similarly, given  $[\mathbf{U}[k]]_{:,1:N_S}$ , we have the corresponding solution to the combiner, denoted by  $(\tilde{\mathbf{W}}_P, \tilde{\mathbf{W}}_B[k] \forall k)$ .

Comparing **Algorithm 1** with **Algorithm 2**, we know that the first algorithm uses the received coupling coefficients as the inputs, while the second uses the singular vectors of the channel as the inputs. The coupling coefficients are commonly used for channel estimation [17], [22], but in this paper we use them to directly implement the hybrid beamforming on both sides. As a result, we can get rid of the overhead of channel estimation.

To clearly present the difference in throughput in the simulation results, the calculated throughput values are normalized to the throughput achieved by fully digital beamforming given by

$$I_{DBF} = \frac{1}{K} \sum_{k=0}^{K-1} \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \gamma [\mathbf{\Sigma}^2[k]]_{n_s, n_s} \right), \quad (40)$$

where  $\gamma = \frac{1}{N_S \sigma_n^2}$  and the diagonal entries of  $\mathbf{\Sigma}^2[k]$  are the eigenvalues of  $\tilde{\mathbf{H}}[k] \tilde{\mathbf{H}}^H[k]$ . The data rates achieved by (40) used for the normalization from SNR = -20 dB to 30 dB (step by 5 dB) are: {0.05, 0.14, 0.41, 1.03, 2.17, 3.77, 5.79, 8.17, 10.91, 13.95, 17.13} in bit/s/Hz. In what follows, a complete analysis with respect to three different codebooks, whose coherence values are 0, 0.12, and 0.99, is provided<sup>4</sup>.

### A. Orthogonal codebooks

Assume that the codebooks  $\mathcal{F}$  and  $\mathcal{W}$  have the same number  $N_F = N_W = 32$  of candidates for the analog beamforming vectors. When  $\mathcal{F}$  and  $\mathcal{W}$  are orthogonal codebooks, the 32 candidates for the steering spatial frequency are equally distributed in the spatial frequency domain and the corresponding steering angles are:  $\left\{ \frac{180^\circ}{\pi} \cdot \sin^{-1} \left( \frac{(n_f-16)}{16} \right), n_f = 1, \dots, 32 \right\}$  [35].

In Fig. 6, we evaluate the achievable data rates with  $M = 2, 3, 4, 5$  initially selected analog beam pairs in the proposed method, and more details of these curves are stated as follows:

- $I_{Pro}(Eig, M = 2, 3, 4, 5)$  is calculated by

$$I_{Pro} = \frac{1}{K} \sum_{k=0}^{K-1} I(\hat{\mathbf{F}}_P, \hat{\mathbf{W}}_P, \hat{\mathbf{F}}_B[k], \hat{\mathbf{W}}_B[k]), \quad (41)$$

where  $(\hat{\mathbf{F}}_P, \hat{\mathbf{W}}_P, \hat{\mathbf{F}}_B[k], \hat{\mathbf{W}}_B[k])$  is the output of **Algorithm 1** with the beam selection criterion  $f(\hat{\mathbf{H}}_{E,i}[k]) = \sum_{n_s=1}^{N_S} \log_2(1 + \gamma [\hat{\mathbf{\Sigma}}_{E,i}^2[k]]_{n_s, n_s})$  in Step 6 in the algorithm. In the phase of initial analog beam selection, we reserve  $M = 2, 3, 4, 5$  initially selected analog beam pairs.

- $I_{Pro}(Eig, NF, M = 2, 3, 4, 5)$  is calculated by the same way as  $I_{Pro}(Eig, M = 2, 3, 4, 5)$  but with noise-free observations. That is to say, the inputs of **Algorithm**

<sup>4</sup>The coherence of a codebook  $\mathcal{F}$  is defined as  $\max_{i < j} \frac{|\tilde{\mathbf{f}}_i^H \tilde{\mathbf{f}}_j|}{\|\tilde{\mathbf{f}}_i\|_2 \|\tilde{\mathbf{f}}_j\|_2}$  [34].

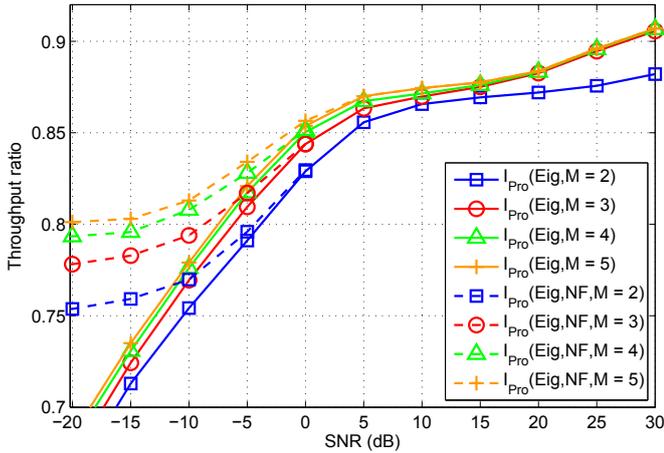


Figure 6. Achievable throughput, normalized to  $I_{DBF}$ , by the proposed methods with the orthogonal codebook and  $M = 2, 3, 4, 5$  initially selected analog beam pairs.

1,  $\{y_{n_w, n_f}[k] \forall n_w, n_f, k\}$ , do not take into account the noise effect.

As shown in Fig. 6, in the low SNR regime, the beam selection performance more or less suffers from the noise effect. When  $\text{SNR} > 0$  dB, the noisy observations are reliable enough to achieve almost the same throughput as that by using the noise-free observations. In addition, when  $M > 5$ , the achievable data rates are almost the same as the curves with  $M = 5$  (although they are not shown in the figure)<sup>5</sup>. For the sake of low complexity, it is not necessary to take more than  $M = 5$  candidates into account because most observations are dominated by noise signals except for those corresponding to the selected analog beam candidates.

Comparisons between the proposed and reference methods are shown in Fig. 7. To better compare our approach with the reference method, we choose the curves  $I_{Pro}(Eig, NF, M = 2, 3, 4, 5)$  in Fig. 6, whose inputs are noise-free observations. Furthermore, the reference curve denoted by  $I_{Ref}$  is calculated by

$$I_{Ref} = \frac{1}{K} \sum_{k=0}^{K-1} I(\check{\mathbf{F}}_P, \check{\mathbf{W}}_P, \check{\mathbf{F}}_B[k], \check{\mathbf{W}}_B[k]), \quad (42)$$

where  $(\check{\mathbf{F}}_P, \check{\mathbf{W}}_P, \check{\mathbf{F}}_B[k], \check{\mathbf{W}}_B[k])$  is obtained from **Algorithm 2** with the inputs  $\{\mathbf{V}[k] \forall k\}$  and  $\{\mathbf{U}[k] \forall k\}$ . In the figure, we can find that the curves  $I_{Pro}(Eig, NF, M = 3, 4, 5)$  achieve higher data rates than  $I_{Ref}$ . Although these two methods use different ways to implement the hybrid beamforming, we try an explanation based on some assumptions. Assume that these two schemes find the same  $N_{RF}$  analog beam pairs (i.e.,  $\hat{\mathbf{F}}_P = \check{\mathbf{F}}_P$  and  $\hat{\mathbf{W}}_P = \check{\mathbf{W}}_P$ ), which means that they have the same effective channel  $\hat{\mathbf{H}}_E[k] = \hat{\mathbf{W}}_P^H \mathbf{H}[k] \hat{\mathbf{F}}_P = \check{\mathbf{W}}_P^H \mathbf{H}[k] \check{\mathbf{F}}_P$ . In this case, **Algorithm 1** uses the SVD of  $\hat{\mathbf{H}}_E[k]$  to find the solution of digital beamforming matrices. From [26], we know that this solution is optimal. In contrast,

<sup>5</sup>The number “ $M - N_{RF}$ ” can be interpreted as a degree of diversity. As it is well known from other diversity techniques, the gain in performance decays more or less quickly with the increasing degrees of diversity.

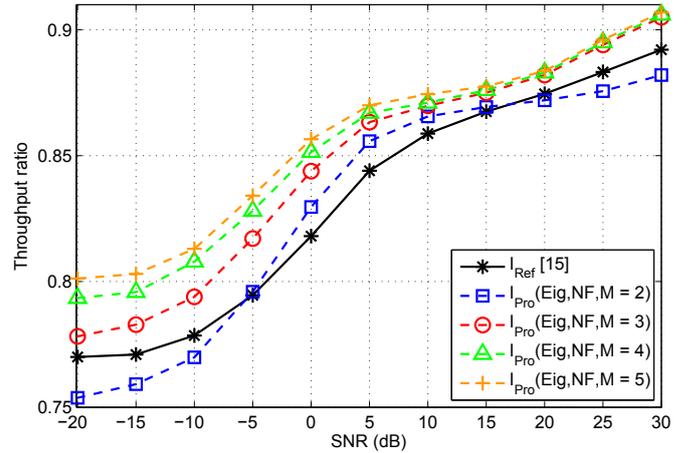


Figure 7. Comparisons between the reference method and the proposed approach with the noise-free (NF) observations, i.e., the curves  $I_{Pro}(Eig, NF, M = 2, 3, 4, 5)$  in Fig. 6.

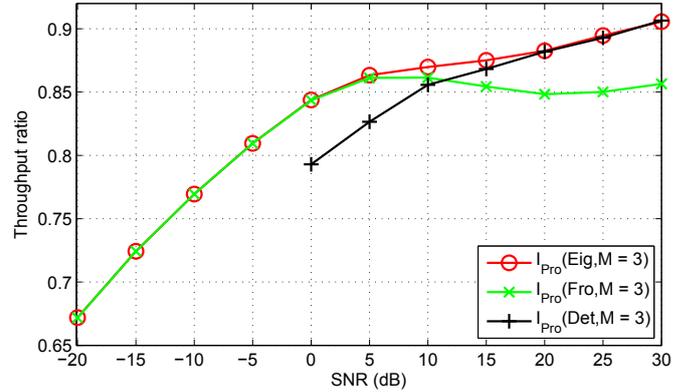


Figure 8. The curve  $I_{Pro}(Eig, M = 3)$  shown in Fig. 6 and its approximations achieved by using two different key parameters.

the digital beamforming in **Algorithm 2** Step 9 uses the least-squares solution, which is sub-optimal. When we reserve more candidates ( $M > N_{RF}$ ), it has a high probability that both algorithms find the same  $N_{RF}$  analog beam pairs. If so, **Algorithm 1** theoretically outperforms **Algorithm 2**.

Next, approximation results of  $I_{Pro}(Eig, M = 3)$  by using the key parameters are shown in Fig. 8, where

- $I_{Pro}(Fro, M = 3)$  and  $I_{Pro}(Det, M = 3)$  use beam selection criteria  $f(\hat{\mathbf{H}}_{E,i}[k]) = \|\hat{\mathbf{H}}_{E,i}[k]\|_F^2$  and  $f(\hat{\mathbf{H}}_{E,i}[k]) = |\det(\hat{\mathbf{H}}_{E,i}[k])|^2$ , respectively, in **Algorithm 1** Step 6.

When  $\text{SNR} < 5$  dB,  $I_{Pro}(Fro, M = 3)$  achieves almost the same throughput as  $I_{Pro}(Eig, M = 3)$ . When  $5 \text{ dB} < \text{SNR} < 20$  dB, both key parameters cannot perfectly yield the same data rates as  $I_{Pro}(Eig, M = 3)$ , but the relative loss amounts to at most a few percentages. From Fig. 6, Fig. 7, and Fig. 8, we can find that if the system operates in the SNR range of 0 to 5 dB, the Frobenius norm of the estimated effective channel works pretty well.

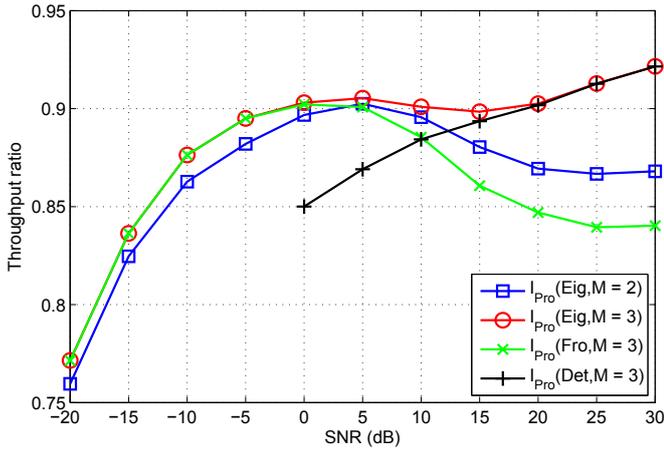


Figure 9. Achievable throughput, normalized to  $I_{DBF}$ , by the proposed approach with the weakly coherent codebook and comparisons between  $I_{Pro}(Eig, M = 3)$  and its approximations.

### B. Non-orthogonal codebooks

Now assume that  $\mathcal{F}$  and  $\mathcal{W}$  are non-orthogonal codebooks. As mentioned in Section V, if the columns of  $\mathcal{F}$  or  $\mathcal{W}$  are not mutually orthogonal, some highly correlated columns (e.g.,  $\frac{|\hat{\mathbf{f}}_i^H \hat{\mathbf{f}}_j|}{\|\hat{\mathbf{f}}_i\|_2 \|\hat{\mathbf{f}}_j\|_2} = 0.99, i \neq j$ ) may make the effective noise level unacceptably large, and the estimated effective channel becomes unreliable accordingly. Here we use two non-orthogonal codebooks to characterize the noise effect:

- The first non-orthogonal codebook has  $N_F = N_W = 36$  columns and the corresponding 36 steering angles are:  $\left\{ \frac{180^\circ}{\pi} \cdot \sin^{-1} \left( \frac{(n_f - 18)}{18} \right), n_f = 1, \dots, 36 \right\}$ . The coherence of the codebook is 0.12 that implies a weakly coherent codebook.
- The second non-orthogonal codebook has larger coherence than the first one. It has  $N_F = N_W = 32$  columns and the corresponding 32 steering angles are:  $\left\{ -90^\circ + \frac{180^\circ \cdot n_f}{N_F}, n_f = 1, \dots, 32 \right\}$ . This codebook design leads to the coherence of 0.99 that implies a strongly coherent codebook.

In Fig. 9, when using the weakly coherent codebook (coherence = 0.12) at the transmitter and receiver, the throughput shown in the curve  $I_{Pro}(Eig, M = 3)$  can be further improved compared with  $I_{Pro}(Eig, M = 2)$ , which means that, in (29), the effect of the correlated columns of the weakly coherent codebook on  $\hat{\mathbf{H}}_{E,i}[k]$  is minor. Also, the approximations of  $I_{Pro}(Eig, M = 3)$  by using the key parameters shown in the curves  $I_{Pro}(Fro, M = 3)$  at low SNR and  $I_{Pro}(Det, M = 3)$  at high SNR are quite accurate and only with small differences in the range between 0 dB and 20 dB. Compared with the results shown in Fig. 8 with the orthogonal codebook, although the approximations in the case of the weakly coherent codebook become slightly worse, the achievable throughput overall becomes better since there are four additional candidates in the weakly coherent codebook.

With the other non-orthogonal codebook whose coherence is 0.99, see Fig. 10, unfortunately the throughput degrades with

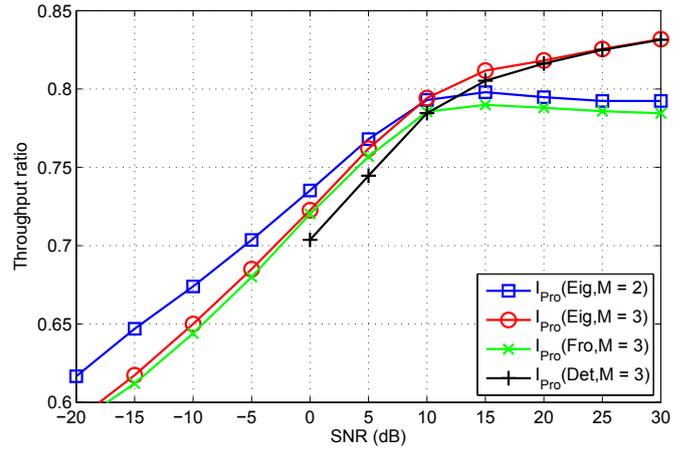


Figure 10. Achievable throughput, normalized to  $I_{DBF}$ , by the proposed approach with the strongly coherent codebook and comparisons between  $I_{Pro}(Eig, M = 3)$  and its approximations.

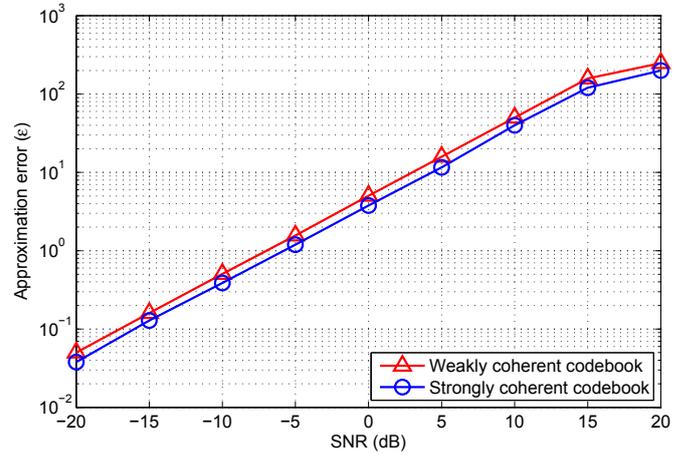


Figure 11. Approximation error between  $I_{Pro}(Eig, NF, M = 3)$  and  $I_{Pro}(Fro, NF, M = 3)$ .

the increasing  $M$  when  $\text{SNR} < 10$  dB. From (30), it is clear that, when we select some highly correlated columns of the codebook, that the variances of the elements of  $\text{vec}(\mathbf{Z}_{E,i}[k])$  are increased leads to unreliable estimates of the effective channel, especially in the low SNR regime. With larger  $M$ , it has a higher probability of selecting these unreliable estimates.

In Fig. 9 and Fig. 10, when  $\text{SNR} > 5$  dB, the gap between  $I_{Pro}(Eig, M = 3)$  and  $I_{Pro}(Fro, M = 3)$  is obvious, but it is not clear that either the approximation error between (21) and (22) or the effective noise dominates the performance loss. To this end, we further provide Fig. 11 to show the approximation error, which are calculated by the following steps. First, using **Algorithm 1** with the noise-free observations and selection criterion  $f(\mathbf{H}_{E,i}[k]) = \sum_{n_s=1}^{N_S} \log_2(1 + \gamma [\boldsymbol{\Sigma}_{E,i}^2[k]]_{n_s, n_s})$  to obtain  $(\hat{\mathbf{F}}_P, \hat{\mathbf{W}}_P)$  and the corresponding noise-free effective channel matrix written as

$$\mathbf{H}_{E,i}[k] = (\hat{\mathbf{W}}_P^H \hat{\mathbf{W}}_P)^{-0.5} \hat{\mathbf{W}}_P^H \mathbf{H}[k] \hat{\mathbf{F}}_P (\hat{\mathbf{F}}_P^H \hat{\mathbf{F}}_P)^{-0.5}. \quad (43)$$

Then using  $\mathbf{H}_{E,i}[k]$  to calculate the approximation error

between (21) and (22) given by

$$\begin{aligned} \epsilon &= \frac{1}{K} \left| \mathbb{E} \left[ \underbrace{\sum_{k=0}^{K-1} \sum_{n_s=1}^{N_s} \log_2 \left( 1 + \gamma \left[ \Sigma_{E,\hat{i}}^2[k] \right]_{n_s, n_s} \right)}_{(21)} \right] \right. \\ &\quad \left. - \mathbb{E} \left[ \underbrace{\sum_{k=0}^{K-1} \gamma \left\| \mathbf{H}_{E,\hat{i}}[k] \right\|_F^2}_{(22)} \right] \right| \quad (44) \\ &\approx \frac{\gamma}{K} \left( \frac{1}{\ln(2)} - 1 \right) \cdot \mathbb{E} \left[ \sum_{k=0}^{K-1} \left\| \mathbf{H}_{E,\hat{i}}[k] \right\|_F^2 \right], \end{aligned}$$

where the diagonal elements of  $\Sigma_{E,\hat{i}}^2$  are the eigenvalues of  $\mathbf{H}_{E,\hat{i}} \mathbf{H}_{E,\hat{i}}^H$ . Repeating the steps with the two non-orthogonal codebooks yields the approximation errors shown in Fig. 11. The approximation error is proportional to the SNR value (or  $\gamma$ ). As a result, at high SNR, the gaps between  $I_{Pro}(Eig, M = 3)$  and  $I_{Pro}(Fro, M = 3)$  in Fig. 9 and Fig. 10 become obvious. At low SNR, because the approximation error between (21) and (22) is small, we can also use the analysis results of (22) in Section V-A and Section V-B to explain the noise effect on  $I_{Pro}(Eig, M = 2, 3)$  with the orthogonal and non-orthogonal codebooks. Moreover, the approximation error between  $I_{Pro}(Eig, M = 3)$  and  $I_{Pro}(Fro, M = 3)$  with the strongly coherent codebook at low SNR in Fig. 10 seems larger than the error with the weakly coherent codebook at low SNR in Fig. 9. Nevertheless, from Fig. 11, we can find that without considering the noise effect, the approximation error with the strongly coherent codebook is even smaller than the results with the other codebook.

In the numerical results, we use these three different types of codebooks to analyze the performance of the proposed method. Generally speaking, the proposed method works well by using orthogonal and weakly coherent codebooks.

## VII. CONCLUSION

This paper presents a novel strategy for the implementation of hybrid beamforming. It shows that hybrid beamforming matrices at the transmitter and receiver can be easily implemented based on the received coupling coefficients so that channel estimation and singular value decomposition are unnecessary. The idea behind this approach is simple: efficiently evaluating the key parameters of the hybrid beamforming gain, such as the Frobenius norm of the effective channel or the absolute value of the determinant of the effective channel. Since the key parameters are functions of the effective channel matrix, which has a much smaller size typically, it is not difficult to try a (small) set of possible alternatives to find a reasonable approximation of the optimal hybrid beamforming matrices. The improvement achieved by additional alternatives can be viewed as a diversity effect that is available from multiple different pairs of array patterns. Moreover, the effective channel matrix can be obtained from the estimated coupling coefficients. This avoids acquiring *explicit* channel estimates and knowledge of the specific angles of propagation paths. Instead, *implicit*

channel knowledge in the sense of which beam pairs produce the strongest coupling between transmitter and receiver is sufficient. Compared with hybrid beamforming methods based on the explicit CSI, the proposed algorithm facilitates the low-complexity hybrid beamforming implementation and should be even robust to deviations from certain desired ideal beam patterns, which is a weak point of all methods based on compressed sensing techniques.

## APPENDIX A

### DERIVATION OF $\bar{\mathbf{F}}_{B,i}[k]$ AND $\bar{\mathbf{W}}_{B,i}[k]$

In the problem (15), if there exists  $\mathbf{W}_B[k]$  such that

$$\mathbf{W}_B^H[k] \bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w} \mathbf{W}_B[k] = \mathbf{I}_{N_s}, \quad (45)$$

one can define a matrix  $\mathbf{Q}_W[k] = (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{0.5} \mathbf{W}_B[k]$ , which is equivalent to  $\mathbf{W}_B[k] = (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} \mathbf{Q}_W[k]$ . Replacing  $\mathbf{W}_B[k]$  in (45) by  $(\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} \mathbf{Q}_W[k]$  leads to  $\mathbf{Q}_W^H[k] \mathbf{Q}_W[k] = \mathbf{I}_{N_s}$  so that the columns of  $\mathbf{Q}_W[k]$  are mutually orthogonal. Similarly, if there exists  $\mathbf{F}_B[k]$  that satisfies the other power constraint at the transmitter, we can define another matrix  $\mathbf{Q}_F[k] = (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{0.5} \mathbf{F}_B[k]$ , which is equivalent to  $\mathbf{F}_B[k] = (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{-0.5} \mathbf{Q}_F[k]$  [19], [20].

Given  $\mathbf{W}_B[k] = (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} \mathbf{Q}_W[k]$  and  $\mathbf{F}_B[k] = (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{-0.5} \mathbf{Q}_F[k]$ , the objective function of the problem therefore becomes

$$\begin{aligned} &I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \mathbf{F}_B[k], \mathbf{W}_B[k]) \\ &= \log_2 \det \left( \mathbf{I}_{N_s} + \frac{1}{\sigma_n^2} \mathbf{Q}_W^H[k] \mathbf{H}_{E,i}[k] \mathbf{Q}_F[k] \mathbf{R}_s \mathbf{Q}_F^H[k] \mathbf{H}_{E,i}^H[k] \mathbf{Q}_W[k] \right), \end{aligned}$$

where  $\mathbf{H}_{E,i}[k]$ ,  $i = (i_f - 1)I_W + i_w$ , is the effective channel defined as

$$\mathbf{H}_{E,i}[k] \triangleq (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} \bar{\mathbf{W}}_{P,i_w}^H \mathbf{H}[k] \bar{\mathbf{F}}_{P,i_f} (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{-0.5}.$$

Let the SVD of  $\mathbf{H}_{E,i}[k]$  be

$$\mathbf{H}_{E,i}[k] \stackrel{\text{SVD}}{=} \mathbf{U}_{E,i}[k] \Sigma_{E,i}[k] \mathbf{V}_{E,i}^H[k],$$

the throughput at subcarrier  $k$  is bounded by

$$\begin{aligned} &I(\bar{\mathbf{F}}_{P,i_f}, \bar{\mathbf{W}}_{P,i_w}, \mathbf{F}_B[k], \mathbf{W}_B[k]) \\ &\leq \sum_{n_s=1}^{N_s} \log_2 \left( 1 + \frac{1}{\sigma_n^2} \left[ \Sigma_{E,i}^2[k] \right]_{n_s, n_s} \left[ \mathbf{R}_s \right]_{n_s, n_s} \right) \end{aligned}$$

with equality if  $\mathbf{Q}_W[k] = [\mathbf{U}_{E,i}[k]]_{:,1:N_s}$ , where the columns of  $\mathbf{Q}_W[k]$  are mutually orthogonal as required, and  $\mathbf{Q}_F[k] = [\mathbf{V}_{E,i}[k]]_{:,1:N_s}$ , which satisfies the condition  $\text{tr}(\bar{\mathbf{F}}_{P,i_f} \mathbf{F}_B[k] \mathbf{R}_s \mathbf{F}_B^H[k] \bar{\mathbf{F}}_{P,i_f}^H) = \text{tr}(\mathbf{R}_s)$  when  $\mathbf{F}_B[k] = (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{-0.5} \mathbf{Q}_F[k]$ . As a result, the solution to the maximization problem is given by

$$\begin{aligned} \bar{\mathbf{F}}_{B,i}[k] &= (\bar{\mathbf{F}}_{P,i_f}^H \bar{\mathbf{F}}_{P,i_f})^{-0.5} [\mathbf{V}_{E,i}[k]]_{:,1:N_s}, \\ \bar{\mathbf{W}}_{B,i}[k] &= (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} [\mathbf{U}_{E,i}[k]]_{:,1:N_s}. \end{aligned}$$

## APPENDIX B

DERIVATION OF THE COVARIANCE MATRIX OF  $\text{VEC}(\mathbf{Z}_{E,i}[k])$

Let us repeat (29) that  $\mathbf{Z}_{E,i}[k] = (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-0.5} \mathbf{Z}[k] (\overline{\mathbf{F}}_{P,i_f}^H \overline{\mathbf{F}}_{P,i_f})^{-0.5}$ , where the elements of  $\mathbf{Z}[k]$  have the same normal distribution with mean zero and variance  $\sigma_n^2$ . To find the covariance between the elements of  $\mathbf{Z}_{E,i}[k]$ , we vectorize  $\mathbf{Z}_{E,i}[k]$  as

$$\begin{aligned} \text{vec}(\mathbf{Z}_{E,i}[k]) &= \left( \left( (\overline{\mathbf{F}}_{P,i_f}^H \overline{\mathbf{F}}_{P,i_f})^{-0.5} \right)^T \otimes (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-0.5} \right) \text{vec}(\mathbf{Z}[k]) \\ &= \left( (\overline{\mathbf{F}}_{P,i_f}^T \overline{\mathbf{F}}_{P,i_f}^*)^{-0.5} \otimes (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-0.5} \right) \text{vec}(\mathbf{Z}[k]), \end{aligned}$$

and the covariance matrix of  $\text{vec}(\mathbf{Z}_{E,i}[k])$  is given by (46).

## APPENDIX C

DERIVATION OF  $E[U]$  AND  $\text{VAR}(U)$  FOR NON-ORTHOGONAL CODEBOOKS

By the definition of  $U$  in (32), one has the mean and variance of  $U$  given by

$$\begin{aligned} E[U] &= E \left[ \|\mathbf{Z}_{E,i}[k]\|_F^2 \right] \\ &= E \left[ \text{tr} \left( \text{vec}(\mathbf{Z}_{E,i}[k]) \text{vec}(\mathbf{Z}_{E,i}[k])^H \right) \right] \\ &= \sigma_n^2 \text{tr} \left( (\overline{\mathbf{F}}_{P,i_f}^T \overline{\mathbf{F}}_{P,i_f}^*)^{-1} \right) \text{tr} \left( (\overline{\mathbf{W}}_{P,i_w}^H \overline{\mathbf{W}}_{P,i_w})^{-1} \right) \end{aligned}$$

and  $\text{Var}(U)$  in (47).

## REFERENCES

- [1] T. Rappaport, R. Heath, R. Daniels, and J. Murdock, *Millimeter Wave Wireless Communications*. Prentice Hall, 2014.
- [2] T. A. Thomas, H. C. Nguyen, G. R. MacCartney, and T. S. Rappaport, "3D mmwave channel model proposal," in *IEEE Veh. Technol. Conf. (VTC Fall)*, Vancouver, BC, Canada, Sep. 2014, pp. 1–6.
- [3] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
- [4] K. Haneda, S. L. H. Nguyen, J. Järveläinen, and J. Putkonen, "Estimating the omni-directional pathloss from directional channel sounding," in *European Conf. on Antennas and Propagation (EuCAP)*, Davos, Switzerland, Apr. 2016, pp. 1–5.
- [5] 3GPP TR 38.900 V14.3.1, "Study on channel model for frequency spectrum above 6 GHz (Release 14)," Tech. Rep., 2017.
- [6] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [7] V. Frascolla and et al., "Challenges and opportunities for millimeter-wave mobile access standardisation," in *IEEE Globecom Workshops (GC Wkshps)*, Austin, TX, USA, Dec. 2014, pp. 553–558.
- [8] J. Liberti and T. Rappaport, *Smart antennas for wireless communications: IS-95 and third generation CDMA applications*. Prentice Hall, 1999.
- [9] A. Hajimiri, H. Hashemi, A. Natarajan, X. Guan, and A. Komijani, "Integrated phased array systems in silicon," *Proc. IEEE*, vol. 93, no. 9, pp. 1637–1655, Sep. 2005.
- [10] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Wiley, 2002.
- [11] 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.
- [12] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [13] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5g cellular communications: theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
- [14] S. Han, C. I. I. Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.
- [15] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *IEEE Int. Conf. on Commun. (ICC)*, Ottawa, ON, Canada, Jun. 2012, pp. 3724–3729.
- [16] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "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.
- [17] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.
- [18] H. L. Chiang, T. Kadur, W. Rave, and G. Fettweis, "Low-complexity spatial channel estimation and hybrid beamforming for millimeter wave links," in *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Valencia, Spain, Sep. 2016, pp. 942–948.
- [19] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
- [20] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for ofdm-based large-scale mimo systems," in *IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, Edinburgh, UK, Jul. 2016, pp. 1–7.
- [21] —, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.
- [22] H. L. Chiang, W. Rave, T. Kadur, and G. Fettweis, "Full rank spatial channel estimation at millimeter wave systems," in *Int. Symp. on Wireless Commun. Syst. (ISWCS)*, Poznan, Poland, Sep. 2016, pp. 42–48.
- [23] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
- [24] Z. Xiao, P. Xia, and X. G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Jan. 2017.
- [25] K. Venugopal, A. Alkhateeb, R. W. Heath, and N. G. Prelcic, "Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 6493–6497.
- [26] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. on Telecommun.*, vol. 10, pp. 585–595, 1999.
- [27] H. L. Chiang, W. Rave, T. Kadur, and G. Fettweis, "A low-complexity beamforming method by orthogonal codebooks for millimeter wave links," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 3375–3379.
- [28] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.
- [29] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.
- [30] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [31] 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.
- [32] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.
- [33] H. L. Chiang, W. Rave, T. Kadur, and G. Fettweis, "Hybrid beamforming strategy for wideband millimeter wave channel models," in *Int. ITG Workshop on Smart Antennas (WSA)*, Berlin, Germany, Sep. 2017, pp. 1–7.
- [34] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, pp. 59–73, Jul. 2011.
- [35] H. L. Chiang, T. Kadur, and G. Fettweis, "Analyses of orthogonal and non-orthogonal steering vectors at millimeter wave systems," in *IEEE Int. Symp. on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Coimbra, Portugal, Jun. 2016, pp. 1–6.

$$\begin{aligned}
\mathbf{R}_{Z_{E,i}} &\triangleq \mathbb{E} [\text{vec}(\mathbf{Z}_{E,i}[k])\text{vec}(\mathbf{Z}_{E,i}[k])^H] \\
&= \mathbb{E} \left[ \left( (\bar{\mathbf{F}}_{P,i_f}^T \bar{\mathbf{F}}_{P,i_f}^*)^{-0.5} \otimes (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} \right) \text{vec}(\mathbf{Z}[k])\text{vec}(\mathbf{Z}[k])^H \left( (\bar{\mathbf{F}}_{P,i_f}^T \bar{\mathbf{F}}_{P,i_f}^*)^{-0.5} \otimes (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-0.5} \right)^H \right] \\
&= \sigma_n^2 \left( (\bar{\mathbf{F}}_{P,i_f}^T \bar{\mathbf{F}}_{P,i_f}^*)^{-1} \otimes (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-1} \right)
\end{aligned} \tag{46}$$

$$\begin{aligned}
\text{Var}(U) &= \mathbb{E} [U^2] - \mathbb{E} [U]^2 = \mathbb{E} [\|\mathbf{Z}_{E,i}[k]\|_F^4] - \mathbb{E} [U]^2 \\
&= \mathbb{E} \left[ \text{tr} \left( \underbrace{\left( (\bar{\mathbf{F}}_{P,i_f}^T \bar{\mathbf{F}}_{P,i_f}^*)^{-1} \otimes (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-1} \right) \otimes \left( (\bar{\mathbf{F}}_{P,i_f}^T \bar{\mathbf{F}}_{P,i_f}^*)^{-1} \otimes (\bar{\mathbf{W}}_{P,i_w}^H \bar{\mathbf{W}}_{P,i_w})^{-1} \right)}_{\Psi \in \mathbb{C}^{N_{RF}^4 \times N_{RF}^4}} \right) \right. \\
&\quad \cdot \left. \left( (\text{vec}(\mathbf{Z}[k])\text{vec}(\mathbf{Z}[k])^H) \otimes (\text{vec}(\mathbf{Z}[k])\text{vec}(\mathbf{Z}[k])^H) \right) \right) - \mathbb{E} [U]^2 \\
&= \text{tr} (\Psi \cdot \mathbb{E} [(\text{vec}(\mathbf{Z}[k])\text{vec}(\mathbf{Z}[k])^H) \otimes (\text{vec}(\mathbf{Z}[k])\text{vec}(\mathbf{Z}[k])^H)]) - \mathbb{E} [U]^2 \\
&= \text{tr} \left( \Psi \cdot \mathbb{E} \left[ \underbrace{(\text{vec}(\mathbf{Z}[k]) \otimes \text{vec}(\mathbf{Z}[k]))}_{\mathbf{z}_V[k] \in \mathbb{C}^{N_{RF}^4 \times 1}} (\text{vec}(\mathbf{Z}[k]) \otimes \text{vec}(\mathbf{Z}[k]))^H}_{\mathbf{z}_V^H[k]} \right] \right) - \mathbb{E} [U]^2 \\
&= \text{tr} (\Psi \cdot \mathbf{R}_{z_V}) - \mathbb{E} [U]^2
\end{aligned} \tag{47}$$



channel estimation, and MIMO techniques.

**Hsiao-Lan Chiang** received the M.S. degree from the Institute of Electronics Engineering, National Chiao Tung University, Taiwan, in 2007. From 2007 to 2014, she was an algorithm development engineer with MediaTek, Taiwan, and Industrial Technology Research Institute, Taiwan. She is now pursuing the Ph.D. degree at the Institute of Communications Engineering, Technische Universität Dresden, Germany. Her research interests lie in the areas of digital signal processing and wireless communications, including millimeter wave beamforming,



awards. In Dresden he has spun-out ten startups, and setup funded projects of more than EUR 1/3 billion volume. He has helped organizing IEEE conferences, most notably as TPC Chair of IEEE ICC 2009, IEEE TTM 2012, and General Chair of VTC Spring 2013. He remains active within IEEE.

**Gerhard Fettweis** earned his Ph.D. under H. Meyr's supervision from RWTH Aachen in 1990. After one year at IBM Research in San Jose, CA, he moved to TCSI Inc., Berkeley, CA. Since 1994, he has been a Vodafone Chair Professor at TU Dresden, Germany, with currently 20 companies from Asia/Europe/US sponsoring his research on wireless transmission and chip design. He coordinates 2 DFG centers at TU Dresden, cfAED and HAEC. Gerhard is an IEEE Fellow, member of acadtech. He has an honorary doctorate from TU Tampere and has received multiple



communications, such as spatial multiplexing, the compensation of RF impairments in OFDM, UWB communications, and aspects of estimation and detection theory.

**Wolfgang Rave** is a senior lecturer at the Vodafone Chair Mobile Communications Systems of Technische Universität (TU) Dresden. He received the diploma and Ph.D. degrees in materials science for electrical engineering from the University of Erlangen-Nürnberg as well as a habilitation degree in electrical engineering from TU Dresden for work on numerical micromagnetics. Since joining the Vodafone Chair in 1999, his research has focused on the application of signal processing techniques to different problems in the context of mobile communications,



**Tobias Kadur** received his Dipl.-Ing. degree in electrical engineering from Technische Universität Dresden, Dresden, Germany, in 2013. During his studies, he focused on wireless communications. Since 2014, he has been a Research Associate at the Vodafone Chair at TU Dresden. He is now pursuing his Ph.D. degree. His research interests focus on millimeter wave communications and beamforming.