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. To implement the hybrid beamforming at the transceiver, a common solution is to decouple the transmitter and receiver functions and then optimize them separately based on given channel state information. However, many references ignore the complexity of the high-dimensional channel estimation problem or the effort for the subsequent step of computing the singular value decomposition of a large 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 a coupling of the channel and all pairs of possible analog beamforming vectors at the transmitter and receiver, and it is obtained by receiving the transmitted pilots. Based on the received pilots, different effective channel matrices are constructed in the sense that the matrix elements are essentially the coupling coefficients. Instead of calculating mutual information, we use the Frobenius norm of the effective channel as a key parameter of the hybrid beamforming gain. As a result, the complicated hybrid beamforming problem becomes a much simpler one: it amounts to trying different sets of the large-power received pilots to construct multiple alternatives for the effective channel matrix. Then, the set yielding the largest Frobenius norm provides the solution to the hybrid beamforming problem.

Index Terms

millimeter wave, analog beam selection, frequency-selective fading channel, implicit channel state information, key parameter of the hybrid beamforming gain.

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 (mm-wave) for future wireless communications taking advantage of the enormous amount of available spectrum at frequencies > 6 GHz [1]. Measurements of

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.
mm-wave channel characteristics presented in [2]-[5] show that the path loss in such environment is very severe. In order to improve capacity and service quality, mm-wave small-cell deployment together with beamforming for large antenna arrays is seen as a promising approach [6][7]. When systems operate at mm-wave 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].

Although both the analog and digital beamforming use the same word *beamforming*, only the former has a specific geometrical meaning in the sense of directing or collecting energy towards specific angles using the so-called array pattern of an antenna array. In contrast, the latter has rather an algebraic meaning, that is, a coefficient matrix. According to the functions of analog and digital beamforming, the hybrid beamforming can be regarded as first converting a high-dimensional over-the-air channel matrix $H$ in the spatial domain into an effective channel $H_E$ of significantly smaller dimensions $^1$ in the selected beam space by the analog beamforming. Then, in the selected beam space, one can further design the coefficients defined in the digital beamforming according to some optimality criteria, e.g., to maximize mutual information.

Unquestionably, it is intractable to deal with the hybrid beamforming at the transmitter and receiver simultaneously because there are four unknown matrices (two analog and two digital beamforming matrices) that are needed to address the surrounding environment appropriately. To avoid dealing with all these issues at the same time, one can simplify the problem by initially assuming that the channel matrix is known. Then the problem of hybrid beamforming on both sides (i.e., finding the precoder and combiner) can be solved by utilizing the singular value decomposition (SVD) of the channel matrix [12],[15]-[17], where 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 ones become necessary to handle multi-carrier modulation in frequency-selective fading channels. An alternative to decouple the transceiver is based on the assumption that either the transmitter or receiver employs fully digital beamforming in order to facilitate low-complexity hybrid beamforming based on given channel state information (CSI) [18][19]. Unfortunately, the overhead of some preliminary works, such as the channel estimation for large-scale antenna arrays [20]-[22], make the previously proposed solutions difficult to obtain.

A feasible alternative that decouples the four unknown matrices is therefore introduced as follows. Generally speaking, given candidates of analog beamforming vectors selected from the given codebooks on both sides, finding the optimal solution to the digital beamforming problem is trivial in the sense that it is almost equivalent to the

$^1$The dimensions of $H_E$ are determined by the number of available RF chains on both sides; a more detailed description of this will be given in Section II.
conventional fully digital beamforming apart from different power constraints [23]. As a result, the critical issue of hybrid beamforming is definitely the analog beam selection. The work in [24] explains why the analog beam selection based on the power of the received pilot can be equivalent to that by the orthogonal matching pursuit (OMP) algorithm [25]. This happens when the analog beamforming vectors are selected from orthogonal codebooks (intuitively such codebooks acts as complete dictionaries in terms of compressed sensing techniques). However, the result of the analog beam selection technique based on the received power can be further improved because the corresponding effective channel $H_E$ is not necessarily well-conditioned [26][27]. In simpler words, the factor dominating the performance of hybrid beamforming is the singular values of $H_E$ rather than the received power.

A simple cure for the problem is that one can reserve a few more candidates of the analog beamforming vectors with respect to the large received power levels. Then, the subset of these candidates yielding 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 a key parameter 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 or the square of the determinant of the effective channel matrix $H_E$, where the effective channel can be regarded as a coupling of the channel and analog beamforming on both sides. Such coupling coefficients can be obtained by receiving the transmitted pilots. In other words, the observations (i.e., the received pilots) used for the analog beam selection can also be used to construct the effective channel matrix $H_E$. Accordingly, the collected received pilots yielding the maximum value of the key parameter gives us the necessary information to optimally select the analog beamforming vectors.

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

1) Most hybrid beamforming methods in the literature are implemented based on the explicit CSI and ignore the complexity of high-dimensional channel estimation or SVD. To avoid the channel estimation, this paper therefore presents a method that uses implicit CSI (i.e., the received pilots or coupling coefficients) to directly construct the hybrid beamforming at the transmitter and receiver.

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 constructed simultaneously based on the received coupling coefficients with reasonable complexity.

3) Compared with existing approaches to hybrid beamforming, we propose a different perspective on problem
solving by using the Frobenius norm of the effective channel as the key parameter to determine the hybrid beamforming gain. This significantly alleviates the complexity of the problem. Although codebook training effort is still non-negligible, it appears unavoidable in any case or for any type of approach.

This rest of the paper is organized as follows: Section II describes the models for the system and the mm-wave frequency-selective fading channel. Section III defines the objective function and formulates precisely the optimization problem of hybrid beamforming. Based on the objective function, a solution is outlined that leads to the implicit-CSI-based hybrid beamforming algorithm presented in Section IV. A theoretical analysis in terms of statistical 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. $a$ is a scalar, $a$ is a column vector, and $A$ is a matrix. $a_n$ denotes the $n^{th}$ column vector of $A$; $[A]_{n,n}$ denotes the $n^{th}$ diagonal element of $A$; $[A]_{1:N}$ denotes the first $N$ column vectors of $A$; $[A]_{1:N,1:N}$ denotes the $N \times N$ submatrix extracted from the upper-left corner of $A$. $A^*$, $A^H$, and $A^T$ denote the complex conjugate, Hermitian transpose and transpose of $A$. $\|A\|_F$ denotes the Frobenius norm of $A$. $\det(A)$ denotes the determinant of $A$; $\text{vec}(A)$ denotes vectorization of $A$. $A \otimes B$ denotes the Kronecker product of $A$ and $B$. $I_N$ and $0_{N \times M}$ denote the $N \times N$ identity and $N \times M$ zero matrices.

II. SYSTEM MODEL

Assume that the system has a transmitter with a uniform linear antenna array (ULA) consisting 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 $F_P F_B[k]$ at each subcarrier $k = 1, \cdots, K$ consisting of a analog beamforming matrix $F_P \in \mathbb{C}^{N_T \times N_{RF}}$ (implemented in passband as part of the RF front end) and a digital beamforming matrix $F_B[k] \in \mathbb{C}^{N_{RF} \times N_S}$ (in baseband). The value $K$ specifies the number of subcarriers in one OFDM symbol as well as the FFT/IFFT size, and $N_{RF}$ denotes the number of available RF chains, assumed to be equal at both the transmitter and receiver for simplicity. Costs and power consumption impose hardware constraint on $F_P$ such that the degrees of freedom of $F_P$ are much less than that of $F_B[k]$. Specifically, first, $F_P$
should be a constant matrix within (at least) one OFDM symbol which requires a certain coherence time of the channel. Second, the entries of \( F_P \) have equal magnitude because analog beamformers are typically implemented by delay elements in the RF front end. The \( N_{RF} \) analog beamforming vectors (also known as steering vectors) of \( F_P \) are selected from a predefined codebook \( F = \{ \hat{f}_{n_j} \in \mathbb{C}^{N_T \times 1}, n_f = 1, \cdots, N_F \} \), where \( N_F \leq N_T \). Particularly, \( \hat{f}_{n_j} \) can be represented as \[8\]

\[
\hat{f}_{n_j} = \frac{1}{\sqrt{N_T}} \begin{bmatrix} 1, e^{j \frac{2 \pi}{\lambda_0} \sin(\phi_{T,n_f}) \Delta_d}, \cdots, e^{j \frac{2 \pi}{\lambda_0} \sin(\phi_{T,n_f}) \cdot (N_T - 1) \Delta_d} \end{bmatrix}^T,
\]

where \( \phi_{T,n_f} \) stands for the \( n_f \)th candidate of 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 \( W_P W_B[k] \) has a similar structure as the precoder, where \( W_P \in \mathbb{C}^{N_R \times N_{RF}} \) and \( W_B[k] \in \mathbb{C}^{N_{RF} \times N_S} \) are the analog and digital beamforming matrices, respectively. Also, the columns of \( W_P \) are selected from a codebook \( \mathcal{W} = \{ \hat{w}_{n_w} \in \mathbb{C}^{N_R \times 1}, n_w = 1, \cdots, N_W \} \), where \( N_W \leq N_R \) and the columns 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 \( H[k] \in \mathbb{C}^{N_R \times N_T} \), the received signal \( r[k] \in \mathbb{C}^{N_S \times 1} \) at subcarrier \( k \) can be written as

\[
r[k] = \sqrt{\rho} \cdot W_B^H[k] W_B H[k] F_P F_B[k] s[k] + W_B^H[k] W_B H[k] x[k] + z[k],
\]

where \( \rho \) stands for the average received power containing the transmit power, transmit antenna gain, receive antenna gain, and path loss, \( s[k] \in \mathbb{C}^{N_S \times 1} \) is the transmitted pilot or data whose covariance matrix \( R_s = \text{E}[s[k] s^H[k]] \) is a diagonal matrix, and \( n[k] \in \mathbb{C}^{N_R \times 1} \) is an \( N_R \)-dimensional independent and identically distributed (i.i.d.) jointly circularly symmetric complex Gaussian random vector, i.e., \( n[k] \sim \mathcal{CN}(0_{N_R \times 1}, \sigma_n^2 I_{N_R}) \). Furthermore, the precoded transmitted signal \( x[k] \) and combined noise vector \( z[k] \) are enforced to satisfy the following two conditions: constant transmit power at each subcarrier\(^2\) and i.i.d. combined noise vector \( z[k] \)

\[
\begin{align*}
\text{tr}(R_z) &= \text{tr} \left( F_P F_B[k] R_s F_B^H[k] F_P^H \right) = \text{tr}(R_s), \\
R_z &= \sigma_z^2 W_B^H[k] W_B W_B[k] = \sigma_z^2 I_{N_S},
\end{align*}
\]

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

Properties of mm-wave channel have been widely studied recently, and simulation models have been developed accordingly. The most comprehensive one can be found in [5] released in 2017. Based on the references [5][28], a simplified cluster-based frequency-selective fading channel has \( C \) clusters and \( R \) rays of each cluster. At subcarrier

\(^2\)We consider a stricter condition that the constant power is allocated per subcarrier instead of per OFDM symbol.
Fig. 2. An example of the cluster-based channel model with the inter-cluster AoD and AoA ($\phi_{D,c}$ and $\phi_{A,c}$) and cluster angular spread ($c_D \Delta_r$ and $c_A \Delta_r$).

$k$, the channel matrix can be written as

$$H[k] = \sqrt{\frac{1}{P}} \sum_{c=1}^{C} \sum_{r=1}^{R} \alpha_{c,r} \cdot e^{-j2\pi k l_{c,r}/K} \cdot a_A(\phi_{A,c,r})a_D(\phi_{D,c,r})^H,$$  \hspace{1cm} (4)

where the channel characteristics are given by the following parameters:

- $\alpha_{c,r} \in \mathbb{R}_{>0}$ describes the inter- and intra-cluster power, where the mean of difference between light-of-sight (LoS) and NLoS cluster power values is 20 dB, and $P = \sum_{c=1}^{C} \sum_{r=1}^{R} |\alpha_{c,r}|^2$ corresponds to the total power of all $C \cdot R$ paths.\(^3\)

- Frequency-selective properties of the channel are specified in terms of quantized-normalized delays (i.e. delay indices measured in unit of the sampling interval) $l_{c,r} = \lfloor \tau_{c,r} F_S \rfloor \in \mathbb{R}_{\geq 0}$, where $\tau_{c,r}$ and $F_S$ stand for the path delay and sampling rate.

- Intra-cluster angle of departure (AoD) $\phi_{D,c,r}$ of ray $r$ in cluster $c$ is characterized by its mean $\phi_{D,c}$, cluster-wise root mean square (RMS) angle spread of departure angles $c_D$, and offset angle $\Delta_r$ for ray $r$, see Fig. 2,

$$\phi_{D,c,r} = \phi_{D,c} + c_D \Delta_r,$$ \hspace{1cm} (5)

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

- The departure array response vector $a_D(\phi_{D,c,r})$ with unit euclidean norm has entries of equal magnitude and is a function of AoD only (due to the assumption that $\Delta_d = \lambda_0/2$) shown as

$$a_D(\phi_{D,c,r}) = \frac{1}{\sqrt{NT}} \left[ 1, e^{j \frac{2\pi}{\lambda_0} \sin(\phi_{D,c,r}) \Delta_d}, \ldots, e^{j \frac{2\pi}{\lambda_0} \sin(\phi_{D,c,r})(NT-1) \Delta_d} \right]^T.$$ \hspace{1cm} (6)

The corresponding arrival array response vector $a_A(\phi_{A,c,r})$ has the same form as (6).

\(^3\)The path loss value, which is influenced by environment (urban or rural), propagation medium (dry or moist air), distance between the transmitter and receiver, and height and location of antennas, is described in the averaged received power $\rho$ in (2).
III. PROBLEM STATEMENT

The objective of the precoder matrices $F_P F_B[k] \forall k$ and the associated combiner matrices $W_P W_B[k] \forall k$ is to achieve maximum throughput across all the subcarrier subject to the constraints on $F_P$, $W_P$, $F_B[k]$, and $W_B[k] \forall k$, i.e. we seek matrices that solve

$$\max_{F_P, W_P, (F_B[k], W_B[k]) \forall k} \sum_{k=0}^{K-1} I(F_P, W_P, F_B[k], W_B[k]),$$

subject to

$$f_{P,n_f} \in \mathcal{F}, w_{P,n_f} \in \mathcal{W} \forall n_f,$$

$$W_B^H[k] W_P^H W_B[k] = I_{N_S} \forall k,$$

$$\text{tr} (F_P F_B[k] R_s F_B^H[k] F_P^H) = \text{tr}(R_s) \forall k,$$

where the last two constraints are the consequences of (3) and the throughput for subcarrier $k$ is defined as [18]

$$I(F_P, W_P, F_B[k], W_B[k]) = \log_2 \det \left( I_{N_S} + \rho R_s^{-1} (W_B^H[k] W_P^H H[k] F_P F_B[k]) R_s (W_B^H[k] W_P^H H[k] F_P F_B[k])^H \right).$$

The solution to (7) are denoted as $(F_{P,Opt}, W_{P,Opt}, (F_{B,Opt}[k], W_{B,Opt[k]}) \forall k)$.

The original problem shown in (7) is too difficult to solve; therefore, we reformulate it in an alternative way as follows. Given two sets $\mathcal{I}_F$ and $\mathcal{I}_W$ containing the candidates of $F_P$ and $W_P$, the achievable data rate in (7) is greater than or equal to

$$\max_{F_P \in \mathcal{I}_F, W_P \in \mathcal{I}_W} \sum_{k=0}^{K-1} I(F_P, W_P, F_B[k], W_B[k]),$$

subject to

$$W_B^H[k] W_P^H W_B[k] = I_{N_S} \forall k,$$

$$\text{tr} (F_P F_B[k] R_s F_B^H[k] F_P^H) = \text{tr}(R_s) \forall k.$$ 

These two versions of the hybrid beamforming problems will have the same data rates if $\mathcal{I}_F$ and $\mathcal{I}_W$ contain $F_{P,Opt}$ and $W_{P,Opt}$, respectively.

The reformulated problem in (9) becomes simpler because, given $F_P$ and $W_P$, the inner problem is equivalent to conventional fully digital beamforming designs subject to different power constraints have already provided solutions to obtain the local maximum throughput $I_{LM}(F_P, W_P)$ [10][23][29]. In simpler words, the critical issue of hybrid beamforming is to solve the outer problem by an additional maximization over all members of $\mathcal{I}_F$ and $\mathcal{I}_W$. Therefore, the motivation is to find $\mathcal{I}_F$ and $\mathcal{I}_W$, which ideally contain $F_{P,Opt}$ and $W_{P,Opt}$ and perhaps few other candidates, and then select a pair of $(F_P, W_P)$ from $\mathcal{I}_F$ and $\mathcal{I}_W$ leads to the maximum throughput.
IV. HYBRID BEAMFORMING ALGORITHM BASED ON IMPLICIT CSI

A. Initial analog beam selection

To begin with, let us see how to obtain the sets $\mathcal{I}_F$ and $\mathcal{I}_W$ in (9) from the given codebooks $F$ and $W$, which can be orthogonal or non-orthogonal ones. We call this step “initial analog beam selection”. Since the hardware-constrained analog beamforming matrices $F_P$ and $W_P$ cannot be replaced by the identity ones, one has to train all columns of the codebooks $F = \{\tilde{f}_{n_f}, n_f = 1, \cdots, N_F\}$ and $W = \{\tilde{w}_{n_w}, n_w = 1, \cdots, N_W\}$ by transmitting known pilot signals $\{s[k]\}_{k=0}^{K-1}$, satisfying $|s[k]|^2 = 1 \forall k$. For this purpose we use one OFDM symbol to train every beam pair. Then, the observation used for the analog beam selection at subcarrier $k$ for a specific beam pair, $\tilde{w}_{n_w}$ and $\tilde{f}_{n_f}$, is obtained by correlating the received signal with its transmitted pilot

$$ y_{n_w,n_f}[k] = \frac{s^*[k]}{|s[k]|^2} \left( \sqrt{\rho} \tilde{w}_{n_w}^H \mathbf{H}[k] \tilde{f}_{n_f} s[k] + \tilde{w}_{n_w}^H \mathbf{n}[k] \right) $$

$$ = \sqrt{\rho} \tilde{w}_{n_w}^H \mathbf{H}[k] \tilde{f}_{n_f} s[k] + z_{n_w,n_f}[k].$$

Similar observations become available on all the subcarriers $k = 0, \cdots, K$, and the combined noise $z_{n_w,n_f}[k] \sim \mathcal{CN}(0, \sigma_n^2)$ still has a Gaussian distribution with mean zero and variance $\sigma_n^2$ due to rotational symmetry of the original noise probability density function (PDF). The observation $y_{n_w,n_f}[k]$ can also be viewed as a coupling coefficient corresponds to a pair of single beamforming vector on both sides over the channel.

Borrowing the idea from our previous work in [24][30], it shows that if the columns of $F$ and $W$ are orthogonal, the sum of the power of $K$ observations in one OFDM symbol is the minimum-variance unbiased (MVU) estimate of the received power level and can be directly used for analog beam selection, so that the $M$ (assume $M \geq N_{RF}$) analog beam pairs can be selected individually and sequentially by the following selection criterion

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

where $m = 1, \cdots, M$ and $F' = \{\hat{f}_{P,n}, n = 1, \cdots, m-1\}$ and $W' = \{\hat{w}_{P,n}, n = 1, \cdots, m-1\}$ are the sets containing already selected analog beamforming vectors during the iterations 1 to $m-1$. We assume that $M \geq N_{RF}$ for the reason that the first $N_{RF}$ selected analog beams according to the sorted magnitude of $\sum_{k=0}^{K-1} |y_{n_w,n_f}[k]|^2$ may not always lead to the optimal solution. Specifically, we do not consider the effect of the digital beamforming during the analog beam selection phase. To achieve the optimal solution, one has to further take into account the linear combination of $N_{RF}$ analog beamforming vectors with coefficients defined in the digital beamforming.

After selecting $M$ analog beam pairs, we define two sets $\mathcal{I}_F$ and $\mathcal{I}_W$ containing all the combinations of any $N_{RF}$ distinct analog beamforming vectors selected from $\{\hat{f}_{P,m}, m = 1, \cdots, M\}$ and $\{\hat{w}_{P,m}, m = 1, \cdots, M\}$,
The achievable data rate by the received power of the coupling coefficients is 11 bit/s/Hz.

The achievable data rate by the linear combination of two analog beamforming vectors is 11.5 bit/s/Hz.

Fig. 3. A typical example of analog beam selection by two different approaches. In the simplified two-path channel model, the AoDs are \{5°, 30°\}, the AoAs are \{5°, −15°\}, and the difference in attenuation between path one and two amounts to 11.6 dB.

respectively, which can be written as

\[ I_F = \{ F_{P,if}, if = 1, \cdots, I_F \} \]

\[ I_W = \{ W_{P,iw}, iw = 1, \cdots, I_W \} \]

where the cardinality \( I_F = I_W = (M \over N_{RF}) \) of both sets is given by the binomial coefficient. The notation \( F_{P,if} \) and \( W_{P,iw} \) denote the \( i_f^{th} \) and \( i_w^{th} \) candidates of the analog beamforming matrices \( F_P \) and \( W_P \). When \( M \) becomes large, there is a high probability that \( I_F \) and \( I_W \) contains the global optimum solution \( F_{P,Opt} \) and \( W_{P,Opt} \). One schematic example is given as follows.

**Schematic example:** consider a system with \( N_T = N_R = 8 \) antenna elements, codebook sizes \( N_F = N_W = 8 \). Let us assume orthogonal codebooks for simplicity with steering angles given by \{−90°(or 90°), −48.59°, −30°, −14.48°, 0°, 14.48°, 30°, 48.59°\} and \( N_{RF} = 2 \) available RF chains to transmit \( N_S = 2 \) data streams at SNR = 0 dB.

The channel realization as depicted in Fig. 3 has two paths. First, in Fig. 3(a), it shows two analog beam pairs selected according to the sorted power level estimates that steer towards these two paths (highlighted in red). Before the 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 according to the criterion expressed in (11), more options with digital beamforming can be explored. In this example, with \( M = 4 \), we have \( I_F = I_W = (M \over N_{RF}) = (4 \over 2) = 6 \) members in both \( I_F \) and \( I_W \), see Fig. 4. We can enumerate them explicitly as

\[ I_F = \{ F_{P,if}, if = 1, \cdots, 6 \} \]

\[ I_W = \{ W_{P,iw}, iw = 1, \cdots, 6 \} \]

For instance, \( F_{P,1} = [\tilde{f}_5, \tilde{f}_6] \) and \( W_{P,1} = [\tilde{w}_3, \tilde{w}_4] \). Therefore one can try all 36 pairs of the members of \( I_F \) and \( I_W \) to determine the optimal weightings defined in the digital beamforming that can yield the maximum throughput, which will be detailed in the following sections. 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, as shown in Fig. 3(b) with the beams highlighted in blue, the two analog beam pairs steering to the dominant path leads to higher spectral efficiency. However, which beamforming strategy leads to higher throughput with a higher probability it 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 candidates of \( F_P \) and \( W_P \), and the objective is to rapidly find the optimal solution to the joint hybrid beamforming problem. Before going into the detail of our proposed scheme, let us review the relationship between the analog and digital beamforming. Given one particular choice \( (F_{P,i_f}, W_{P,i_w}) \) 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 throughput \( I_{LM}(F_{P,i_f}, W_{P,i_w}) \) as defined in (9) with the objective function shown as

\[
\max_{(F_{B,i}, W_{B,i})} \sum_{k=0}^{K-1} I(F_{P,i_f}, W_{P,i_w}, F_{B[k]}, W_{B[k]})
\]

\[
= \sum_{k=0}^{K-1} \max_{F_{B,i}, W_{B,i}} I(F_{P,i_f}, W_{P,i_w}, F_{B[k]}, W_{B[k]}),
\]

where \( I(F_{P,i_f}, W_{P,i_w}, F_{B[k]}, W_{B[k]}) \) is given by (8) with \( F_B = F_{P,i_f} \) and \( W_B = W_{P,i_w} \). As a consequence, for subcarrier \( k \) the digital beamforming problem can be formulated as throughput maximization problem subject to the defined power constraints, which can be stated as

\[
(F_{B,i[k]}, W_{B,i[k]}) = \arg \max_{F_{B[i]}, W_{B[i]}} I(F_{P,i_f}, W_{P,i_w}, F_{B[k]}, W_{B[k]})
\]

s.t.

\[
W_B^H[k] = I_{N_S},
\]

\[
\text{tr} \left( F_{P,i_f} F_{B[k]} R_s F_{P,i_f}^H[k] F_{B[k]}^H \right) = \text{tr}(R_s).
\]

where \( i = (i_f - 1)I_W + i_w \) is an index specifying the combined elements from \( \mathcal{I}_F \) and \( \mathcal{I}_W \). To proceed, we take advantage of the following mathematical results:
Lemma 1. Given two matrices $Q_W[k] \in \mathbb{C}^{N_{RF} \times N_s}$ and $Q_F[k] \in \mathbb{C}^{N_{RF} \times N_s}$ with orthogonal columns, then

1. The condition $W_B[k]^H W_{P,iw}^H W_{P,iw} W_B[k] = I_{N_s}$ holds if $W_B[k] = (W_{P,iw}^H W_{P,iw})^{-0.5} Q_W[k]$.
2. The condition $\text{tr}(F_{P,i}^H F_B[k] R_s F_{P,i}^H[k]^* F_{P,i}^H) = \text{tr}(R_s)$ holds if $F_B[k] = (F_{P,i}^H F_{P,i}^H)^{-0.5} Q_F[k]$.

Proof: see Appendix A.

Lemma 2. Given $F_{P,i}$, $W_{P,iw}$, and

$$F_B[k] \stackrel{\text{Lemma 1}}{=} (F_{P,i}^H F_{P,i})^{-0.5} Q_F[k],$$

$$W_B[k] \stackrel{\text{Lemma 1}}{=} (W_{P,iw}^H W_{P,iw})^{-0.5} Q_W[k].$$

One has the maximum achievable throughput if $Q_F[k] = [V_{E,i}[k]]_{i=1:N_s}$ and $Q_W[k] = [U_{E,i}[k]]_{i=1:N_s}$, where the columns of $V_{E,i}[k]$ and $U_{E,i}[k]$ are respectively the right and left singular vectors of the effective channel

$$H_{E,i}[k] \triangleq (W_{P,iw}^H W_{P,iw})^{-0.5} W_{P,iw}^H H[k] F_{P,i} (F_{P,i}^H F_{P,i})^{-0.5} \stackrel{\text{SVD}}{=} U_{E,i}[k] \Sigma_{E,i}[k] V_{E,i}^H[k].$$

Proof: see Appendix B.

It is worth noting that the effective channel $H_{E,i}[k] \in \mathbb{C}^{N_{RF} \times N_s}$ corresponds to the effect of the coupling between the analog beamforming matrices $(F_{P,i}, W_{P,iw})$ and the over-the-air channel $H[k]$. Based on Lemma 1 and 2, we can find that given one pair of analog beamforming matrices $(F_{P,i}, W_{P,iw})$ selected from $\mathcal{I}_F$ and $\mathcal{I}_W$, the corresponding optimal digital beamforming matrices shown as follows are not difficult to obtain as mentioned previously

$$F_{B,i}[k] = (F_{P,i}^H F_{P,i})^{-0.5} [V_{E,i}[k]]_{i=1:N_s},$$

$$W_{B,i}[k] = (W_{P,iw}^H W_{P,iw})^{-0.5} [U_{E,i}[k]]_{i=1:N_s}.$$  \hspace{1cm} (16)

The same solution to $F_B[k]$ with different explanations can be found in [18][19] as the constraint on the transmit power is similar. At the receiver, the power constraint on $W_B[k]$ is not necessarily the same which leads to a different solution.

C. Key parameter of hybrid beamforming

As discussed above, given pairs of elements selected from $\mathcal{I}_F$ and $\mathcal{I}_W$ and the corresponding optimal digital beamforming matrices, we can evaluate the local maximum throughput

$$I_{LM}(F_{P,i}, W_{P,iw}) = \sum_{k=0}^{K-1} I(F_{P,i}, W_{P,iw}, F_{B,i}[k], W_{B,i}[k])$$

$$= \sum_{k=0}^{K-1} \sum_{n_s=1}^{N_s} \log_2 \left( 1 + \frac{\rho}{\sigma_n^2} \mathbf{S}_{E,i}[k]_{n_s,n_s} [R_s]_{n_s,n_s} \right),$$

\hspace{1cm} (17)
where (a) follows from Appendix B and the diagonal elements of \( \Sigma_{E,i}[k] \) are the singular values of the effective channel \( H_{E,i}[k] \). Based on the candidate set \( \{ (\mathbf{F}_{P,i}, \mathbf{W}_{P,i}) \mid \forall i_f, \forall i_w \} \), the pair that leads to the maximum throughput provides the approximation of the optimal analog beamforming matrices, denoted as \( (\hat{\mathbf{F}}_P, \hat{\mathbf{W}}_P) \), as well as the solution to the hybrid beamforming problem in (9)

\[
\left( \mathbf{F}_P, \mathbf{W}_P \right) = \arg \max_{\mathbf{F}_{P,i} \in \mathcal{I}_F, \mathbf{W}_{P,i} \in \mathcal{I}_W} I_L(M(\mathbf{F}_{P,i}, \mathbf{W}_{P,i})).
\]  

(18)

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

To reduce the potentially large computational burden, we ask ourselves what are the crucial parameters or indicators that determine the throughput. To find an answer more easily, 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

\[
I(\mathbf{F}_{P,i}, \mathbf{W}_{P,i}, \mathbf{F}_{B,i}[k], \mathbf{W}_{B,i}[k]) = \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \frac{\rho}{N_S \sigma_n^2} \left[ \Sigma_{E,i}[k] \right]_{n_s,n_s} \right) \approx \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \gamma \left[ \Sigma_{E,i}[k] \right]_{n_s,n_s} \right),
\]

(19)

where \( i = (i_f - 1)I_W + i_w \) and \( \gamma = \frac{\rho}{N_S \sigma_n^2} \) is the SNR. To analyze the key parameters, we focus on the high and low SNR regimes. First, in the low SNR regime \( (\gamma \rightarrow 0) \), using the fact that \( \log(1 + \gamma x) \approx \gamma x \) as \( \gamma \rightarrow 0 \), the achievable data rate in (19) can be approximated by

\[
I(\mathbf{F}_{P,i}, \mathbf{W}_{P,i}, \mathbf{F}_{B,i}[k], \mathbf{W}_{B,i}[k]) \approx \gamma \sum_{n_s=1}^{N_S} \left[ \Sigma_{E,i}[k] \right]_{n_s,n_s} \propto \sum_{n_s=1}^{N_S} \left[ \Sigma_{E,i}[k] \right]_{n_s,n_s} \leq \| H_{E,i}[k] \|_F^2.
\]

(20)

with equality in (b) iff \( N_{RF} = N_S \). For the case of \( N_{RF} > N_S \), \( \| H_{E,i}[k] \|_F^2 \) corresponds to the sum of all \( N_{RF} \) (instead of only the \( N_S \) strongest) eigenvalues of \( H_{E,i}[k]H_{E,i}^H[k] \). Assuming that the sum of the weaker \( N_{RF} - N_S \) eigenvalues of \( H_{E,i}[k]H_{E,i}^H[k] \) is small, the approximation of \( \sum_{n_s=1}^{N_S} \left[ \Sigma_{E,i}[k] \right]_{n_s,n_s} \) by \( \| H_{E,i}[k] \|_F^2 \) 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 (19) is approximated by

\[
I \left( \mathbf{F}_{P,i}, \mathbf{W}_{P,iw}, \mathbf{F}_{B,i}[k], \mathbf{W}_{B,i}[k] \right) \approx \sum_{n_s=1}^{N_S} \log_2 \left( \gamma \left[ \Sigma_{E,i}^2[k] \right]_{n_s,n_s} \right)
\]

\[
\propto \prod_{n_s=1}^{N_S} \left[ \Sigma_{E,i}^2[k] \right]_{n_s,n_s}
\]

which again holds with equality in (c) iff \(N_{RF} = N_S\) since

\[
|\text{det}\left( \mathbf{H}_{E,i}[k] \right)|^2 = \prod_{n=1}^{N_{RF}} |\text{det}\left( \mathbf{E}_{E,i}[k] \right)|^2 |\text{det}\left( \mathbf{V}_E^H[k] \right)|^2 = \prod_{n_s=1}^{N_S} \left[ \Sigma_{E,i}^2[k] \right]_{n_s,n_s}.
\]

To summarize these arguments, either the Frobenius norm or the absolute squared determinant of the effective channel matrix \(\mathbf{H}_E\) acts as the key parameter for the system throughput, shown as

\[
I \left( \mathbf{F}_{P,i}, \mathbf{W}_{P,iw}, \mathbf{F}_{B,i}[k], \mathbf{W}_{B,i}[k] \right) \propto f(\mathbf{H}_{E,i}[k]),
\]

where \(f(\mathbf{H}_{E,i}[k])\) denotes one of the two functions with the argument \(\mathbf{H}_{E,i}[k]\) given by

\[
f(\mathbf{H}_{E,i}[k]) = \begin{cases} \|\mathbf{H}_{E,i}[k]\|^2_F, & \gamma \to 0 \\ |\text{det}\left( \mathbf{H}_{E,i}[k] \right)|^2, & \gamma \to \infty \end{cases}
\]

Fortunately, the matrix \(\mathbf{H}_{E,i}[k]\) can be easily obtained from the original observations \(\{y_{n_{w,f}}[k] \forall n_{w,f}, k\}\) [30]. Let us show the effective channel in (15) again and approximate the matrix \(\mathbf{W}_{P,iw}^H \mathbf{H}[k] \mathbf{F}_{P,i}\) by \(\mathbf{Y}_i[k]\)

\[
\mathbf{H}_{E,i}[k] \triangleq (\mathbf{W}_{P,iw}^H \mathbf{W}_{P,iw})^{-0.5} \mathbf{W}_{P,iw}^H \mathbf{H}[k] \mathbf{F}_{P,i} \mathbf{F}_{P,i}^H \mathbf{W}_{P,iw}(\mathbf{F}_{P,i}^H \mathbf{F}_{P,i})^{-0.5}
\]

\[
\approx (\mathbf{W}_{P,iw}^H \mathbf{W}_{P,iw})^{-0.5} \mathbf{Y}_i[k] (\mathbf{F}_{P,i}^H \mathbf{F}_{P,i})^{-0.5}
\]

\[
\triangleq \mathbf{H}_{E,i}[k],
\]

where the elements of \(\mathbf{Y}_i[k]\) can be collected from the observations. For example, if \(\mathbf{F}_{P,i} = [\mathbf{F}_{P,1}, \cdots, \mathbf{F}_{P,N_{RF}}]\) and \(\mathbf{W}_{P,iw} = [\mathbf{W}_{P,1}, \cdots, \mathbf{W}_{P,N_{RF}}]\), where \(\mathbf{F}_{P,n_{f}}\) is the \(n_{f}\) column of \(\mathcal{F}\) and \(\mathbf{W}_{P,n_{f}}\) is the \(n_{f}\) column of \(\mathcal{W}\),

\[
\mathbf{Y}_i[k] = \begin{bmatrix} y_{1,1}[k] & \cdots & y_{1,N_{RF}}[k] \\
\vdots & \ddots & \vdots \\
y_{N_{RF},1}[k] & \cdots & y_{N_{RF},N_{RF}}[k] \end{bmatrix} = \mathbf{W}_{P,iw}^H \mathbf{H}[k] \mathbf{F}_{P,i} + \mathbf{Z}[k].
\]

Therefore, given a pair \((\mathbf{F}_{P,i}, \mathbf{W}_{P,iw})\) selected from \(\mathcal{I}_\mathcal{F}\) and \(\mathcal{I}_\mathcal{W}\), one can rapidly obtain the approximation of
\( \mathbf{H}_{E,i}[k] \), denoted as \( \hat{\mathbf{H}}_{E,i}[k] \).

To conclude, the proposed solution can be stated as: first obtain the candidate sets \( (\mathcal{I}_F, \mathcal{I}_W) \) 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 (18), which can be rewritten as

\[
\begin{align*}
(\hat{i}_f, \hat{i}_w) &= \arg \max_{\mathbf{F}_{P,i_f} \in \mathcal{I}_F, \mathbf{W}_{P,i_w} \in \mathcal{I}_W} I_{LM}(\mathbf{F}_{P,i_f}, \mathbf{W}_{P,i_w}) \\
&\approx \arg \max_{\mathbf{F}_{P,i_f} \in \mathcal{I}_F, \mathbf{W}_{P,i_w} \in \mathcal{I}_W} \sum_{k=0}^{K-1} f(\hat{\mathbf{H}}_{E,i}[k]).
\end{align*}
\]  

(27)

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

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

(28)

where \( \hat{i} = (\hat{i}_f-1)I_W + \hat{i}_w \) and \( \hat{\mathbf{U}}_{E,i}[k] \hat{\mathbf{V}}_{E,i}[k]^H = \text{svd}(\hat{\mathbf{H}}_{E,i}[k]) \). The complete proposed algorithm of the hybrid beamforming implementation based on the implicit CSI (i.e., the coupling coefficients \( \{y_{n_w,n_f}[k], \forall n_w,n_f,k\} \)) is shown in Algorithm 1, where the key parameter for operation in the low SNR regime is used. The advantage of the proposed algorithm is that, even though the number of members of \( \mathcal{I}_F \) and \( \mathcal{I}_W \) is large, the computational overhead is minor because we just need to calculate the Frobenius norm of the effective channel, whose elements can be easily obtain from the observations \( \{y_{n_w,n_f}[k], \forall n_w,n_f,k\} \).

D. Water-filling power allocation

In the previous subsection, we simply assume that the transmit power is equally allocated to \( N_S \) data streams to facilitate the process of finding the key parameter. In this subsection, we consider that both the analog and digital beamforming matrices are selected, then the global maximum throughputs can be further improved by optimizing the power allocation for \( N_S \) data streams according to the effective channel condition \( \mathbf{H}_{E,i}[k] \) at each subcarrier. Let the covariance matrix of the transmitted signal be \( \mathbf{R}_s = \text{diag}(p_1[k], \cdots, p_{N_S}[k]) \), where \( \sum_{n_s=1}^{N_S} p_{n_s}[k] = 1 \), then the optimal power allocation subject to the total power constraint is given by [23]

\[
\hat{p}_{n_s}[k] = \arg \max_{\{p_{n_s}[k]\}_{n_s=1}^{N_S}} \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \frac{p}{\sigma^2_n} \left[ \mathbf{\Sigma}_{E,i}[k] \right]_{n_s,n_s} p_{n_s}[k] \right), \text{ s.t. } \sum_{n_s=1}^{N_S} p_{n_s}[k] = 1.
\]

(29)
Algorithm 1: Hybrid beamforming by the key parameter in the low SNR regime.

Input: \( \{y_{n_w,m}[k], \forall n_w, n_f, k\} \)

Output: \( \hat{F}_P, \hat{W}_P, \{\hat{F}_B[k]\}_{k=0}^{K-1} \), and \( \{\hat{W}_B[k]\}_{k=0}^{K-1} \)

1. **Part I — Initial analog beam selection**

2. \((\hat{f}_{P,m}, \hat{w}_{P,m}) = \arg\max_{\tilde{f}_{n_j}, \tilde{w}_{n_j} \in \mathcal{F}, \mathcal{W}} \sum_{k=0}^{K-1} |y_{n_w,n_j}[k]|^2\), where \( \mathcal{F} = \{\hat{f}_{P,n}, n = 1, \ldots, m-1\} \), and \( \mathcal{W} = \{\hat{w}_{P,m}, n = 1, \ldots, m-1\} \).

3. **Part II — Analog beam selection by the key parameter**

4. **Part III — Corresponding optimal digital beamforming**

5. \( \hat{H}_{E,i}[k] = (\hat{W}_{P,i_w}^H \mathcal{W}_{P,i_w})^{-0.5} \mathbf{Y}_i[k](\mathbf{F}_{P,i_f}^H \mathbf{F}_{P,i_f})^{-0.5} \), where \( i = (i_f-1)I_W + i_w \), \( \mathbf{F}_{P,i_f} \in \mathcal{F}_W, \hat{w}_{P,i_w} \in \mathcal{I}_W \), entries of \( \mathbf{Y}_i[k] \) are collected from \( \{y_{n_w,n_f}[k] \forall n_w, n_f\} \).

6. \((\hat{i}_f, \hat{i}_w) = \arg\max_{\mathbf{F}_{P,i_f} \in \mathcal{F}_W, \hat{w}_{P,i_w} \in \mathcal{I}_W} \sum_{k=0}^{K-1} \| \hat{H}_{E,i}[k] \|^2 \).

7. **Output:** \( \hat{F}_P = \hat{F}_{P,i_f} \), and \( \hat{W}_P = \hat{W}_{P,i_w} \).

Applying the standard technique of using a Lagrange multiplier to solve this problem, we have

\[
\hat{p}_{n_s}[k] = \mu - \frac{\sigma_n^2}{\rho \left[ \sum_{i,j}^2 \tilde{E}_{i,j}[k] \right]_{n_s,n_s}},
\]

where the constant \( \mu \) satisfies the following equality

\[
\sum_{n_s=1}^{N_s} \left( \mu - \frac{\sigma_n^2}{\rho \left[ \sum_{i,j}^2 \tilde{E}_{i,j}[k] \right]_{n_s,n_s}} \right) = 1,
\]

and the notation \((x)^+ := \max(0, x)\) returns \( x \) if \( x > 0 \), otherwise it returns 0.

V. **ANALYSIS OF THE PROPOSED HYBRID BEAMFORMING ALGORITHM**

In the section, we focus on the statistical analysis of using the (squared) Frobenius norm of the effective channel as the key parameter at low SNR. Starting from (25), \( \tilde{H}_{E,i}[k] \) can be expressed as a noisy version of the unknown true effective channel

\[
\tilde{H}_{E,i}[k] = \left( \mathbf{W}_{P,i_w}^H \mathcal{W}_{P,i_w} \right)^{-0.5} \mathbf{Y}_i[k](\mathbf{F}_{P,i_f}^H \mathbf{F}_{P,i_f})^{-0.5} \\
= \left( \mathbf{W}_{P,i_w}^H \mathcal{W}_{P,i_w} \right)^{-0.5} \mathbf{W}_{P,i_w} \mathbf{H}_{E,i}[k] \mathbf{F}_{P,i_f} \mathbf{F}_{P,i_f}^{-0.5} + \left( \mathbf{W}_{P,i_w}^H \mathcal{W}_{P,i_w} \right)^{-0.5} \mathbf{Z}_{E,i}[k](\mathbf{F}_{P,i_f}^H \mathbf{F}_{P,i_f})^{-0.5} \\
= \mathbf{H}_{E,i}[k] + \mathbf{Z}_{E,i}[k],
\]
where the elements of the matrix $Z_{E,i}[k]$ follow the Gaussian distributions and the covariance matrix of $\text{vec}(Z_{E,i}[k])$ is derived in Appendix C,

$$\text{vec}(Z_{E,i}[k]) \sim \mathcal{CN}\left(0_{N_{RF} \times 1}, \sigma_n^2 \left(\left(F_{P,i}^T \bar{F}_{P,i}^*\right)^{-1} \otimes (\bar{W}_{P,i}^H \bar{W}_{P,i})^{-1}\right)\right).$$  \hspace{1cm} (33)

The objective function of the maximization problem in (27) for the low SNR case can be expressed as the sum of a contribution from the true key parameter value plus some noise term

$$\sum_{k=0}^{K-1} \left\| \hat{H}_{E,i}[k] \right\|_F^2 = \sum_{k=0}^{K-1} \left\| H_{E,i}[k] \right\|_F^2 + \sum_{k=0}^{K-1} \left\| Z_{E,i}[k] \right\|_F^2 + 2 \Re \left( \text{tr}(H_{E,i}^H[k] Z_{E,i}[k]) \right)$$  \hspace{1cm} (34)

To analyze the noise effect, we introduce the quantities

$$U = \sum_{k=0}^{K-1} \left\| Z_{E,i}[k] \right\|_F^2,$$  \hspace{1cm} (35)

$$V = \sum_{k=0}^{K-1} 2 \cdot \Re \left( \text{tr}(H_{E,i}^H[k] Z_{E,i}[k]) \right).$$  \hspace{1cm} (36)

Then, given a constant channel matrix (i.e., one channel realization) and analog beamforming matrices $F_{P,i}$ and $\bar{W}_{P,i}$, we can pursue the analysis of $U$ and $V$ for orthogonal and non-orthogonal codebooks as follows.

### A. Orthogonal Codebooks

When the columns of $\bar{F}_{P,i}$ and $\bar{W}_{P,i}$ are orthogonal, from (33) we know that the elements of $\text{vec}(Z_{E,i}[k])$ are i.i.d. complex Gaussian random variables with mean zero and variance $\sigma_n^2$, $\text{vec}(Z_{E,i}[k]) \sim \mathcal{CN}\left(0_{N_{RF} \times 1}, \sigma_n^2 I_{N_{RF}}\right)$. Therefore, $U$ is the sum of $N_{RF}^2$ i.i.d. absolute squared Gaussian random variables. As a consequence, each term of $U$ individually follows a Gamma distribution with the shape parameter $\frac{1}{2}$ and scale parameter $\sigma_n^2$ as well as $U$ itself whose shape parameter becomes $N_{RF}^2$:

$$U = \sum_{k=0}^{K-1} \sum_{i=1}^{N_{RF}} \sum_{j=1}^{N_{RF}} \Re \left( [Z_{E,i}[k]]_{i,j} \right)^2 + \Im \left( [Z_{E,i}[k]]_{i,j} \right)^2 \sim \Gamma\left(\frac{1}{2}, \sigma_n^2\right),$$  \hspace{1cm} (37)

Also, given one realization of the channel, denoted as $H'[k]$, $V$ has a normal distribution with mean zero and variance $2\sigma_n^2 \left\| \text{vec}(Z_{E,i}[k]) \right\|_F^2$

$$V = \sum_{k=0}^{K-1} 2 \cdot \Re \left( \text{tr}(H_{E,i}^H[k] Z_{E,i}[k]) \right) \sim \mathcal{N}\left(0, 2\sigma_n^2 \sum_{k=0}^{K-1} \left\| Z_{E,i}[k] \right\|_F^2\right),$$  \hspace{1cm} (38)

where $H_{E,i}^H[k] = \bar{W}_{P,i}^H \bar{H}'_{P,i} F_{P,i}$. 
B. Non-orthogonal Codebooks

When the columns of $\mathbf{F}_{P,i,i}$ and $\mathbf{W}_{P,i,w}$ are non-orthogonal, the elements of $\text{vec}(\mathbf{Z}_{E,i}[k])$ in (33) are not i.i.d. any more. In this case, no closed-form expressions for the probability distributions of $U$ and $V$ can be found. Accordingly, we only derive and state the $E[U]$, $\text{Var}(U)$, and $E[V]$ in this section. These are given by (see Appendix D)

$$E[U] = K\sigma_n^2 \cdot \text{tr}\left((\mathbf{F}_{P,i,i}^T \mathbf{F}_{P,i,i})^{-1}\right) \cdot \text{tr}\left((\mathbf{W}_{P,i,w}^H \mathbf{W}_{P,i,w})^{-1}\right),$$

$$\text{Var}(U) = K \cdot \text{tr}(\mathbf{R}_{z_{v'}}) - E[U]^2,$$

where

$$\mathbf{R}_{z_{v'}} = E[(\text{vec}(\mathbf{Z}[k]) \otimes \text{vec}(\mathbf{Z}[k]))(\text{vec}(\mathbf{Z}[k]) \otimes \text{vec}(\mathbf{Z}[k]))^H] \in \mathbb{C}^{N_{RF}^4 \times N_{RF}^4},$$

and $E[V] = 0$ Unfortunately, the variance of $V$, which is a function of $\mathbf{F}_{P,i,i}$ and $\mathbf{W}_{P,i,w}$, is too complicated to have a closed-form expression.

From the analysis results, we can find that the distributions of $U$ and $V$ for non-orthogonal codebooks are both functions of $\mathbf{F}_{P,i,i}$ and $\mathbf{W}_{P,i,w}$, which means that the estimated value of the key parameter, $\|\hat{\mathbf{H}}_{E,i}[k]\|_F^2$ in (34), with respect to different pairs $(\mathbf{F}_{P,i,i}, \mathbf{W}_{P,i,w})$ whose columns are non-orthogonal seriously suffers from the noise with different distributions. As a consequence, the estimate $\|\hat{\mathbf{H}}_{E,i}[k]\|_F^2$ may become unreliable for some pairs. In our numerical results, we will show an example to illustrate the difference in performance between orthogonal and non-orthogonal codebooks.

VI. Simulation Results

A comprehensive frame structure for the new radio in defined [31]; based on this reference, the simulation parameters are defined as follows. The system has $N_T = N_R = 32$ antennas, $N_{RF} = 2$ analog beamforming vectors, $N_S = 2$ data streams, FFT/IFFT size $N_{FT} = 2048$, sampling rate $F_S = 153.6$ MHz, bandwidth of 100 MHz, and carrier frequency of 38 GHz. The cluster-based channel model has $C = 5$ clusters including one LoS and four NLoS clusters, and each cluster has $R = 8$ rays. The noise variance is given by $\sigma_n^2 = \frac{\rho}{N_S}10^{-\gamma/10}$, where $\gamma$ (dB) is the SNR. The water-filling power allocation is applied for all the simulated data rates. Furthermore, to clearly present the simulation results, we also show the ratio of the throughput by the reference and proposed methods to that by the fully digital beamforming. The achievable throughput by the fully digital beamforming is given by

$$\sum_{k=0}^{K-1} \sum_{n_s=1}^{N_s} \log_2 \left(1 + \frac{\rho}{\sigma_n^2} \left[\Sigma^2[k]\right]_{n_s,n_s} p_{n_s}[k]\right),$$

(43)
where the diagonal entries of $\Sigma^2[k]$ are the non-zero eigenvalues of $H[k]H^H[k]$ and $p_n[k]$, $n_s = 1, \cdots, N_S$, are optimized according to the over-the-air channel condition (see Subsection IV-D). In what follows, a complete analysis including two different types of codebooks is given.

A. Orthogonal codebooks

Assume that codebooks $\mathcal{F}$ and $\mathcal{W}$ have the same number of steering vectors, $N_F = N_W = 32$. When the columns of $\mathcal{F}$ and $\mathcal{W}$ are orthogonal, the 32 steering spatial frequency indices are equally distributed in the spatial frequency domain and the corresponding steering angles are: $\{\frac{180^\circ}{\pi}\cdot\sin^{-1}\left(\frac{2\nu}{32}\right), \nu = -15, \cdots, 16\}$, where $\nu$ denotes the spatial frequency index [32][33].
Fig. 7. The normalized achievable throughput by using the Frobenius norm of the effective channel matrix as the key parameter of hybrid beamforming gain (see Algorithm 1) with $M = 2, 3, 4$ initially selected analog beam pairs. The observations $\{y_{n_w,n_f}[k], \forall n_w,n_f,k\}$ used in Algorithm 1 with and without noise are both simulated.

Fig. 5 shows the achievable throughput with $M = 3$ initially selected analog beam pairs by using two reference methods, *Fully DBF* and *Eigenvalue*, and the proposed ones, *Frob. norm* and *Determinant*. Specifically, *Fully DBF* is calculated by (43), *Eigenvalue* is calculated by

$$\sum_{k=0}^{K-1} \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \frac{\rho \sigma_n^2}{\sigma_n^2} \left[ \Sigma_{E,i}^2[k] \right]_{n_s,n_s} p_{n_s}[k] \right), \quad (44)$$

*Frob. norm* is calculated by Algorithm 1, and *Determinant* is calculated by Algorithm 1, where $\left\| \tilde{H}_{E,i}[k] \right\|^2_F$ in Step 6 is replaced by $\left| \det \left( \tilde{H}_{E,i}[k] \right) \right|^2$. In the low SNR regime, theoretically *Frob. norm* has higher data rates than *Determinant*, and ideally it can achieve (almost) the same data rates as *Eigenvalue*. The simulation results shown in Fig. 5 validate the theoretical results. To clearly show the difference in performance between these methods, Fig. 6 shows the normalized achievable throughput, where the throughput by *Eigenvalue, Frob. norm*, and *Determinant* are all normalized to that by *Fully DBF*. When targeting at 80% of the data rate achieved by *Fully DBF*, the required SNR levels of the three methods with $M = 3$ are $-4$ dB for both *Eigenvalue* and *Frob. norm*, and 1 dB for *Determinant*. Since the system operates in the low SNR regime, the achievable data rate by *Eigenvalue* and *Frob. norm* are nearly the same as we expect, while the gap between *Frob. norm* and *Determinant* is obvious because the latter only works in the high SNR regime that is not yet attained here. When SNR is getting larger, the gap is supposed to be smaller.

Previously, the example in Fig. 3 shows that the data rate can be improved by reserving more candidates of

$^4$When $M = N_{RF} = 2$, the three methods *Eigenvalue, Frob. norm*, and *Determinant* have the same throughput since there is only one candidate of $F_P$ and $W_P$ in $I_F$ and $I_{WV}$ respectively.
Fig. 8. The normalized achievable throughput by using the Frobenius norm as the key parameter of hybrid beamforming gain with $M = 2, 3, 4$ initially selected analog beam pairs (see Algorithm 1) and the contaminated observations.

analog beamforming vectors. In Fig. 7, more complete numerical results are given. All results shown in the figure are obtained by using the Frobenius norm of the effective channel matrix as the key parameter (see Algorithm 1). First, we try $M = 2, 3, 4$ initially selected analog beam pairs. When $M = 2$, the $N_{RF} = 2$ analog beam pairs are selected according to the power of the received contaminated coupling coefficients without considering the effect of the linear combination of these two analog beams. If more candidates of analog beamforming vectors are reserved, we can exploit the key parameter to evaluate different alternatives for the corresponding effective channel to find better linear combinations, and the beamforming gain can be further improved as shown in the curves with $M = 2, 3, 4$. When targeting at 80% of the data rate achieved by Fully DBF, the required SNR levels with $M = 2, 3, 4$ are $-1$ dB, $-3.5$ dB, and $-4.5$ dB, respectively. With more than $M = 4$ initially selected analog beam pairs, the achievable data rates are saturated and stay almost the same; this means that the set size $(M_{N_{RF}} = (\frac{4}{2}) = 6$ of $\mathcal{I}_F$ and $\mathcal{I}_W$ is sufficient. Therefore, we do not show results with more than $M = 4$ initially selected beam pairs. Also, cases of the noise-free observations in Algorithm 1 are also considered. Without the noise effect, the performance of both the initial analog beam selection (Algorithm 1, Part I) and analog beam selection by the key parameter (Algorithm 1, Part II) can be improved in the low SNR regime, especially for SNR $< 0$ dB.

B. Non-orthogonal codebooks

Now assume that the columns of $\mathcal{F}$ and $\mathcal{W}$ are non-orthogonal and the $N_F = N_W = 32$ candidates of the steering angles in the codebooks are equally distributed in the angle domain: $\left\{-90^\circ + \frac{180^\circ}{N_F} n_f, n_f = 0, \cdots, N_F - 1 \right\}$.

As discussed in Subsection V-B, the estimated value of the key parameter with respect to different analog beamforming pairs whose columns are non-orthogonal seriously suffers from the noise with different distributions. In Fig. 8, it clearly show the different in performance between the throughput with and without the noise effect.
Fig. 9. The normalized achievable throughput by using the Frobenius norm as the key parameter of hybrid beamforming gain with $M = 2, 3, 4$ initially selected analog beam pairs (see Algorithm 1) and the noise-free observations.

![Non-orthogonal codebook](image)

(a) CDFs of $U_1$ with $(\mathbf{F}_{P,1}, \mathbf{W}_{P_{i,w}}) = (\mathbf{F}_{P,1}, \mathbf{W}_{P,1})$ and $U_2$ with $(\mathbf{F}_{P,2}, \mathbf{W}_{P_{i,w}}) = (\mathbf{F}_{P,2}, \mathbf{W}_{P,2})$.

(b) CDFs of $V_1$ with $(\mathbf{F}_{P,1}, \mathbf{W}_{P_{i,w}}) = (\mathbf{F}_{P,1}, \mathbf{W}_{P,1})$ and $V_2$ with $(\mathbf{F}_{P,2}, \mathbf{W}_{P_{i,w}}) = (\mathbf{F}_{P,2}, \mathbf{W}_{P,2})$.

Fig. 10. The CDFs of the noise terms $U = \sum_{k=0}^{K-1} \| \mathbf{Z}_{E,i}[k] \|_F^2$ and $V = \sum_{k=0}^{K-1} 2 \Re(\text{tr}(\mathbf{H}_{E,i}[k] \mathbf{Z}_{E,i}[k]))$ for the non-orthogonal codebook with one channel realization, where $\mathbf{Z}_{E,i}[k]$ and $\mathbf{H}_{E,i}[k]$ are functions of candidates of the analog beamforming matrices $(\mathbf{F}_{P,i}, \mathbf{W}_{P_{i,w}})$.

With the contaminated observations ($\{y_{n_{w},n_{f},k}, \forall n_{w}, n_{f}, k\}$ in Algorithm 1), the achievable data rates with $M = 4$ initially selected analog beam pairs is even worse than that with $M = 2$ when the SNR < 0 dB. This effect occurs because, in the case of non-orthogonal codebooks, the estimated value of the proposed key parameter $\| \mathbf{H}_{E,i}[k] \|_F^2$ becomes unreliable for some pairs of the elements selected from $\mathcal{I}_F$ and $\mathcal{I}_W$. Therefore, when $M$ becomes large, it has a higher probability that some of the selected $M$ analog beam pairs yield an unreliable estimate of the key parameter. In contrast, when considering the nose-free observations, see Fig. 9, the throughput increases evidently as $M$ is getting larger. More detailed analysis of the noise effect is given as follows.

An example illustrating the problem considers a situation that $\mathcal{I}_F = \{ \mathbf{F}_{P,1}, \mathbf{F}_{P,2} \} = \{ [\tilde{f}_1, \tilde{f}_2], [\tilde{f}_1, \tilde{f}_3] \}$ and $\mathcal{I}_W = \{ \mathbf{W}_{P,1}, \mathbf{W}_{P,2} \} = \{ [\tilde{w}_1, \tilde{w}_2], [\tilde{w}_1, \tilde{w}_3] \}$. As mentioned in Subsection V-B, the distributions of both $U$ and $V$ are functions of the selected candidates of the analog beamforming matrices when $\mathcal{F}$ and $\mathcal{W}$ are non-orthogonal codebooks. Then, given one pair $(\mathbf{F}_{P,1}, \mathbf{W}_{P,1}) = ([\tilde{f}_1, \tilde{f}_2], [\tilde{w}_1, \tilde{w}_2])$, we may denote the corresponding noise terms...
as $U_1$ and $V_1$. Also, given another pair $\left(F_{P,2}, W_{P,2}\right) = \left(\tilde{F}_3, [\tilde{w}_1, \tilde{w}_3]\right)$, let us denote the corresponding noise terms as $U_2$ and $V_2$. Fig. 10(a) and (b) show the results of a numerical evaluation of the corresponding cumulative distribution functions (CDFs) of $U_1$, $U_2$, $V_1$, and $V_2$ with one channel realization, where the possible values of $U_1$ are much larger than that of $U_2$. In this case, the pair $\left(F_{P,1}, W_{P,1}\right)$ has a higher probability to be selected than $\left(F_{P,2}, W_{P,2}\right)$ because of the value of $U_1$ instead of the true value of the key parameter. On the other hand, the effect of $V_1$ and $V_2$ on the key parameter seems minor compared with $U_1$ and $U_2$. The distributions of the noise shown in Fig. 10(a) can explain why the throughput cannot be improved as $M$ is getting larger. For orthogonal codebooks, only the distribution of $V$ is a function of $\left(F_{P,i}, W_{P,i}\right)$, and fortunately the effect of $V$ on the estimate of the key parameter is minor.

Fig. 11 shows the normalized achievable throughput with $M = 3$ initially selected analog beam pairs by three methods \textit{Eigenvalue}, \textit{Frob. norm} and \textit{Determinant}, where the throughput by the three methods are all normalized to that by \textit{Fully DBF} (the curve \textit{Fully DBF} shown in Fig. 5). From Fig. 6 and 11, we can find that no matter what kind of codebooks is used, in the low SNR regime, the achievable data rate by \textit{Eigenvalue} and \textit{Frob. norm} are nearly the same\footnote{The result that \textit{Eigenvalue} and \textit{Frob. norm} achieve almost the same data rates when $M = 3$ means that $\Sigma_{k<4}^2[k]\forall k$ used in \textit{Eigenvalue} also suffer from seriously noise effect as the case in \textit{Frob. norm}. This part is left for future work.}, and they are all greater than that by \textit{Determinant}. When targeting at 80% of the data rate achieved by \textit{Fully DBF}, the required SNR levels of the three methods with $M = 3$ are 6 dB for both \textit{Eigenvalue} and \textit{Frob. norm}, and 7.5 dB for \textit{Determinant}. Compared with the results shown in Fig. 6 with the orthogonal codebook, it requires larger SNR values to achieve the same throughput. One of the reasons is that the orthogonal codebook performs better than the non-orthogonal one when the AoDs and AoAs are uniformly distributed [32]; moreover, the distribution of the effective noise dramatically vary with different analog beam pairs.

Fig. 11. The normalized throughput by the reference method \textit{Eigenvalue} and two proposed methods, \textit{Frob. norm} and \textit{Determinant}, with $M = 3$ initially selected analog beam pairs.
VII. CONCLUSION

This paper presents a novel strategy for the implementation of hybrid beamforming. It shows that the hybrid beamforming matrices at the transmitter and receiver can be easily constructed based on the received coupling coefficients so that the high-dimensional channel estimation and singular value decomposition are unnecessary. The idea behind this approach is simple: efficiently evaluating the key parameter, such as the Frobenius norm of the effective channel matrices, to implement the hybrid beamforming based on the estimates of received power levels. Since the key parameter of the hybrid beamforming gain is a function of the effective channel matrix, which has much lower dimension typically, it is not difficult to try a (small) set of possible alternatives to find a reasonable approximation of the optimal hybrid beamforming. The improvement achieved by the additional alternatives can be viewed as diversity effect that is available from multiple different pairs of array patterns. Moreover, the effective channel matrix can be obtained from the received coupling coefficients. This avoids acquiring the explicit channel estimation and knowledge of the specific angles of the propagation paths. Instead, the implicit channel knowledge that which beam pairs produce the strongest coupling between transmitter and receiver is enough in a sense. Compared with the 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 of the true beam patterns from certain desired ideal patterns, which is a weak point of all methods based on compressed sensing techniques.

VIII. APPENDIX

A. Proof of Lemma 1

Since $\mathbf{W}^H_{P,i} \mathbf{W}_{P,i} \in \mathbb{C}^{N_{RF} \times N_{RF}}$ is a positive definite matrix, it is diagonalizable by a unitary matrix $\mathbf{M} \in \mathbb{C}^{N_{RF} \times N_{RF}}$, where $\mathbf{M}^{-1} \mathbf{M} = \mathbf{I}_{N_{RF}}$. Then, $\mathbf{W}^H_{P,i} \mathbf{W}_{P,i}$ can be represented as $\mathbf{W}^H_{P,i} \mathbf{W}_{P,i} = \mathbf{M} \mathbf{Λ} \mathbf{M}^{-1}$, where $\mathbf{Λ} \in \mathbb{C}^{N_{RF} \times N_{RF}}$ is a diagonal matrix with non-negative real numbers on the diagonal, i.e., the eigenvalues of $\mathbf{W}^H_{P,i} \mathbf{W}_{P,i}$. When $\mathbf{W}_{B,i}[k] = (\mathbf{W}^H_{P,i} \mathbf{W}_{P,i})^{-0.5} \mathbf{Q}_W[k]$, the first constraint becomes

$$\mathbf{W}^H_{B}[k] \mathbf{W}^H_{P,i} \mathbf{W}_{P,i} \mathbf{W}_{B}[k] = \mathbf{Q}^H_W[k](\mathbf{W}^H_{P,i} \mathbf{W}_{P,i})^{-0.5} \mathbf{W}^H_{P,i} \mathbf{W}_{P,i} (\mathbf{W}^H_{P,i} \mathbf{W}_{P,i})^{-0.5} \mathbf{Q}_W[k]$$

$$= \mathbf{Q}^H_W[k] \mathbf{M} \mathbf{Λ}^{-0.5} \mathbf{Λ}^{-0.5} \mathbf{M}^{-1} \mathbf{Q}_W[k]$$

$$= \mathbf{Q}^H_W[k] \mathbf{Q}_W[k]$$

$$= \mathbf{I}_{N_S}.$$  \hspace{1cm} (45)

Similarly, $\mathbf{F}^H_{P,i} \mathbf{F}^{P,i}$ is also diagonalizable by a unitary matrix. Accordingly, when $\mathbf{F}_{B,i}[k] = (\mathbf{F}^H_{P,i} \mathbf{F}^{P,i})^{-0.5} \mathbf{Q}_F[k]$, the second constraint following from the derivation of (45) becomes

$$\text{tr} \left( \mathbf{F}_{P,i} \mathbf{F}^H_{B}[k] \mathbf{Q}_F[k] \mathbf{F}^H_{P,i} \mathbf{F}^{P,i} \mathbf{R}_s \right) = \text{tr} \left( \mathbf{F}^H_B[k] \mathbf{F}^H_{P,i} \mathbf{F}^{P,i} \mathbf{F}^H_B[k] \mathbf{R}_s \right) = \text{tr}(\mathbf{Q}^H_F[k] \mathbf{Q}_F[k] \mathbf{R}_s) = \text{tr}(\mathbf{R}_s).$$  \hspace{1cm} (46)
B. Proof of Lemma 2

Given $F_{P,i,j}$, $W_{P,i,w}$, $F_B[k] = (F_{P,i,w}^H F_{P,i,w})^{-0.5} Q_F[k]$, and $W_B[k] = (W_{P,i,w}^H W_{P,i,w})^{-0.5} Q_W[k]$, the system throughput (8) becomes

$$I(F_{P,i,j}, W_{P,i,w}, F_B[k], W_B[k])$$

$$= \log_2 \det \left( I_{N_S} + \frac{\rho}{\sigma_n^2} Q_W[k] (W_{P,i,w}^H W_{P,i,w})^{-0.5} W_{P,i,w}^H H(k) F_{P,i,j} (F_{P,i,j}^H F_{P,i,j})^{-0.5} Q_F[k] \cdot R_s \cdot Q_W[k] (W_{P,i,w}^H W_{P,i,w})^{-0.5} Q_W[k] \right)$$

$$\leq \sum_{n_s=1}^{N_S} \log_2 \left( 1 + \frac{\rho}{\sigma_n^2} \left| \Sigma_{E,i}[k] \right|_{n_s,n_s} \left| R_s \right|_{n_s,n_s} \right),$$

where (d) follows from $R_s^{-1} = \frac{1}{\sigma_n^2} I_{N_S}$ (see Lemma 1), (e) follows from $H_{E,i}[k] \overset{\text{SVD}}{=} U_{E,i}[k] \Sigma_{E,i}[k] V_{E,i}[k]$, and the equality in (f) holds if $Q_W[k] = [U_{E,i}[k]; 1:N_S]$ and $Q_F[k] = [V_{E,i}[k]; 1:N_S]$ [23][29].

C. Derivation of the covariance matrix of vec($Z_{E,i}[k]$)

The effective noise matrix in (32) is shown as $Z_{E,i}[k] = (W_{P,i,w}^H W_{P,i,w})^{-0.5} Z[k] (F_{P,i,j}^H F_{P,i,j})^{-0.5}$, where the elements of $Z[k]$ has a normal distribution with mean zero and variance $\sigma_n^2$. To see the covariance between the elements of $Z_{E,i}[k]$, we vectorize $Z_{E,i}[k]$ as

$$\text{vec}(Z_{E,i}[k]) = \left( (F_{P,i,j}^H F_{P,i,j})^{-0.5} \right)^T \otimes (W_{P,i,w}^H W_{P,i,w})^{-0.5} \text{vec}(Z[k]) = \left( (F_{P,i,j}^T F_{P,i,j})^{-0.5} \otimes (W_{P,i,w}^H W_{P,i,w})^{-0.5} \right) \text{vec}(Z[k])$$

and define the covariance matrix of vec($Z_{E,i}[k]$) as

$$R_{Z_{E,i}} \triangleq E \left[ \text{vec}(Z_{E,i}[k]) \text{vec}(Z_{E,i}[k])^H \right]$$

$$= E \left[ \left( (F_{P,i,j}^T F_{P,i,j})^{-0.5} \otimes (W_{P,i,w}^H W_{P,i,w})^{-0.5} \right) \text{vec}(Z[k]) \text{vec}(Z[k])^H \left( (F_{P,i,j}^T F_{P,i,j})^{-0.5} \otimes (W_{P,i,w}^H W_{P,i,w})^{-0.5} \right)^H \right]$$

$$= \sigma_n^2 \left( (F_{P,i,j}^T F_{P,i,j})^{-1} \otimes (W_{P,i,w}^H W_{P,i,w})^{-1} \right).$$

(49)

D. Derivation of the $E[U]$, Var($U$), and $E[V]$ for non-orthogonal codebooks

By the definition of $U$ in (35), one has the mean and variance of $U$ shown as

$$E[U] = E \left[ \sum_{k=0}^{K-1} \| Z_{E,i}[k] \|^2 \right] = \sum_{k=0}^{K-1} E \left[ \text{tr} \left( \text{vec}(Z_{E,i}[k]) \text{vec}(Z_{E,i}[k])^H \right) \right] = K \sigma_n^2 \text{tr} \left( (F_{P,i,j}^T F_{P,i,j})^{-1} \right) \text{tr} \left( (W_{P,i,w}^H W_{P,i,w})^{-1} \right)$$

(50)

and
\[
\text{Var}(U) = E[U^2] - E[U]^2
\]
\[
= \sum_{k=0}^{K-1} E\left[\|Z_{E,i}[k]\|^4_F\right] - E[U]^2
\]
\[
= \sum_{k=0}^{K-1} \left[\text{tr}\left(\left(\left(\begin{bmatrix} F_{P,i}^T & F_{P,i}^* \end{bmatrix} - W_{P,i,\omega}^H W_{P,i,\omega}^{-1}\right) \otimes \left(\begin{bmatrix} F_{P,i}^T & F_{P,i}^* \end{bmatrix} - W_{P,i,\omega}^H W_{P,i,\omega}^{-1}\right)\right)_{r \in \mathbb{C}^{N_{RF}^2}}\right)\right] - E[U]^2
\]
\[
= K \cdot \text{tr}\left(\Gamma \cdot E\left[\left(\text{vec}(Z[k])\text{vec}(Z[k])^H\right) \otimes \left(\text{vec}(Z[k])\text{vec}(Z[k])^H\right)\right]\right) - E[U]^2
\]
\[
= K \cdot \text{tr}\left(\Gamma \cdot E\left[\left(\text{vec}(Z[k])\otimes\text{vec}(Z[k])\right)(\text{vec}(Z[k])\otimes\text{vec}(Z[k]))^H\right]_{z_{RF}[k] \in \mathbb{C}^{N_{RF}^2}}\right) - E[U]^2
\]
\[
= K \cdot \text{tr}\left(\Gamma \cdot R_{z_{RF}}\right) - E[U]^2
\]

Also, by the definition of \( V \) in (36) and given one realization of the channel \( H'[k] \), one has the variance of \( V \) shown as
\[
E[V] = E\left[\sum_{k=0}^{K-1} 2 \cdot \Re\left(\text{tr}(H_{E,i}^H[k]Z_{E,i}[k])\right)\right] = 0,
\]
where \( H_{E,i}^H[k] = (W_{P,i,\omega}^H W_{P,i,\omega}^{-1})^{0.5} W_{P,i,\omega}^H H'[k] F_{P,i} F_{P,i}^* (F_{P,i}^H F_{P,i})^{-0.5}. \)

REFERENCES

[1] T. S. Rappaport, R. C. Daniels, and J. N. 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 *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, Sep. 2014.

[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 Transactions on Communications,* 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 *2016 10th European Conference on Antennas and Propagation (EuCAP)*, Apr. 2016.

[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 *2014 IEEE Globecom Workshops*, 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,” *Proceedings of the IEEE,* vol. 93, no. 9, pp. 1637–1655, Sep. 2005.

[10] H. L. Van Trees, *Optimum Array Processing (Part IV).* Wiley, 2004.

[11] X. Zhang, A. F. Molisch, and S.-Y. Kung, “Variable-phase-shift-based rf-baseband codesign for MIMO antenna selection,” *IEEE Transactions on Signal Processing,* 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 Transactions on Wireless Communications, 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 Communications Magazine, 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 Communications Magazine, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[15] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[16] 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.

[17] H. L. Chiang, T. Kadur, W. Rave, and G. Fettweis, “Low-complexity spatial channel estimation and hybrid beamforming for millimeter wave links,” in IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Sep. 2016, pp. 942–948.

[18] A. Alkhateeb and R. W. Heath, “Frequency selective hybrid precoding for limited feedback millimeter wave systems,” IEEE Transactions on Communications, vol. 64, no. 5, pp. 1801–1818, May 2016.

[19] F. Sohrabi and W. Yu, “Hybrid digital and analog beamforming design for large-scale antenna arrays,” IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[20] H. L. Chiang, W. Rave, T. Kadur, and G. Fettweis, “Full rank spatial channel estimation at millimeter wave systems,” in 2016 International Symposium on Wireless Communication Systems (ISWCS), Sep. 2016, pp. 42–48.

[21] 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 Communications Letters, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.

[22] K. Venugopal, A. Alkhateeb, R. W. Heath, and N. G. Prelcic, “Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture,” in 2017 IEEE International Conference on Acoustics, speech and Signal Processing (ICASSP), Mar. 2017.

[23] E. Telatar, “Capacity of multi-antenna gaussian channels,” European Transactions on Telecommunications, vol. 10, pp. 585–595, 1999.

[24] H. L. Chiang, W. Rave, T. Kadur, and G. Fettweis, “A low-complexity beamforming method by orthogonal codebooks for millimeter wave links,” in 2017 IEEE International Conference on Acoustics, speech and Signal Processing (ICASSP), Mar. 2017.

[25] T. T. Cai and L. Wang, “Orthogonal matching pursuit for sparse signal recovery with noise,” IEEE Transactions on Information Theory, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[26] G. H. Golub and C. F. Van Loan, Matrix Computations (3rd Ed.). Johns Hopkins University Press, 1996.

[27] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge University Press, 2005.

[28] IST-WINNER II Deliverable 1.1.2 V1.2, “Winner II channel models,” Tech. Rep., 2007.

[29] H. Bölcskei, D. Gesbert, and A. J. Paulraj, “On the capacity of OFDM-based spatial multiplexing systems,” IEEE Transactions on Communications, vol. 50, no. 2, pp. 225–234, Feb. 2002.

[30] H. L. Chiang, W. Rave, T. Kadur, and G. Fettweis, “Hybrid beamforming strategy for wideband millimeter wave channel models,” in International ITG Workshop on Smart Antennas (WSA), Sep., 2017.

[31] 3GPP TR 38.802 V14.0.0, “Study on new radio access technology physical layer aspects (Release 14),” Tech. Rep., 2017.

[32] H. L. Chiang, T. Kadur, and G. Fettweis, “Analyses of orthogonal and non-orthogonal steering vectors at millimeter wave systems,” in 2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), Jun. 2016.

[33] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, “Hybrid analog-digital precoding revisited under realistic RF modeling,” IEEE Wireless Communications Letters, vol. 5, no. 5, pp. 528–531, Oct. 2016.