Dynamic Subarrays for Hybrid Precoding in Wideband mmWave MIMO Systems

Sungwoo Park, Ahmed Alkhateeb, and Robert W. Heath Jr.

Abstract

Hybrid analog/digital precoding architectures can address the trade-off between achievable spectral efficiency and power consumption in large-scale MIMO systems. This makes it a promising candidate for millimeter wave systems, which require deploying large antenna arrays at both the transmitter and receiver to guarantee sufficient received signal power. Most prior work on hybrid precoding focused on narrowband channels and assumed fully-connected hybrid architectures. MmWave systems, though, are expected to be wideband with frequency selectivity. In this paper, a closed-form solution for fully-connected OFDM-based hybrid analog/digital precoding is developed for frequency selective mmWave systems. This solution is then extended to partially-connected but fixed architectures in which each RF chain is connected to a specific subset of the antennas. The derived solutions give insights into how the hybrid subarray structures should be designed. Based on them, a novel technique that dynamically constructs the hybrid subarrays based on the long-term channel characteristics is developed. Simulation results show that the proposed hybrid precoding solutions achieve spectral efficiencies close to that obtained with fully-digital architectures in wideband mmWave channels. Further, the results indicate that the developed dynamic subarray solution outperforms the fixed hybrid subarray structures in various system and channel conditions.

I. INTRODUCTION

Hybrid analog/digital architectures are efficient transceivers for millimeter wave (mmWave) systems [1]–[6]. These architectures enable a flexible compromise between achieving high spectral efficiency and maintaining low cost and power consumption. Extensive work has been devoted to developing hybrid precoding algorithms to single-user and multi-user mmWave and

Sungwoo Park, Ahmed Alkhateeb, and Robert W. Heath Jr. are with The University of Texas at Austin (Email: spark, aalkhateeb, rheath@utexas.edu).

This work is supported in part by the National Science Foundation under Grant No. 1319556, and by a gift from Huawei Technologies.
massive MIMO systems in the last few years [3], [7]–[13]. Most prior work on hybrid precoding focused on narrowband channels. MmWave systems, however, will likely operate on wideband channels with frequency selectivity [14]–[16]. It is, therefore, important to develop hybrid analog/digital precoding designs for frequency selective mmWave systems.

A. Prior Work

Hybrid architectures divide the processing needed for precoding and combining between analog and digital domains to reduce the number of RF chains [1]–[3], [7]–[13], [17], [18]. In [1], [2], hybrid precoding was first investigated for diversity and multiplexing gains in general MIMO systems. For mmWave large MIMO systems, [3] leveraged the sparse nature of mmWave channels and designed low-complexity hybrid precoding algorithms based on orthogonal matching pursuit. Following [3], the work in [7]–[11] devised hybrid precoding algorithms based on matrix decomposition, alternative minimization, and other techniques, with the objective of achieving spectral efficiencies close to that obtained with fully-digital solutions. The system models in [3], [7]–[11] adopted a fully-connected hybrid architecture, meaning that each RF chain is connected to all the antennas. Extensions to subarray-based hybrid architectures were considered in [12], [13]. The work in [3], [7]–[13], though, assumed a narrowband mmWave channel, with perfect or partial channel knowledge at the transmitter.

Limited work has been done for wideband mmWave hybrid precoding systems. In [17], hybrid beamforming with only a single-stream transmission over MIMO-OFDM systems was considered. The developed solution in [17], though, relied on the exhaustive search over the RF and baseband codebooks, and did not provide specific criteria for the design of these codebooks. In [18], for OFDM-based mmWave hybrid precoding systems, the optimal baseband precoders for a given RF codebook were obtained, and efficient codebooks were designed. The work in [18], however, did not exploit the channel correlation over the adjacent subcarriers to reduce the precoder design complexity. Further, the work in [17], [18] considered only the fully-connected hybrid architecture, which consumes more power consumption compared to the subarray structure [12], [13], which connects each RF chain to only a subset of the antennas.

B. Contribution

In this paper, we develop hybrid precoding designs for wideband mmWave large MIMO systems. The contributions of this paper are summarized as follows.
We develop a near-optimal closed-form solution for fully-connected and partially-connected hybrid analog/digital precoding in OFDM-based wideband mmWave systems. In our design, we assume fully-digital receivers and adopt a relaxation of the original mutual information maximization problem. For the relaxed problem, we obtain the optimal baseband and RF precoders. The developed solution has exactly the same spectral efficiency as the unconstrained fully-digital solution if the number of channel paths is less than the number of RF chains. Therefore, thanks to the sparse nature of the mmWave channel, the proposed hybrid precoding with a small number of RF chains can achieve a spectral efficiency near to that obtained with the unconstrained fully-digitalized baseband precoding. Further, the developed closed-form solution provides insights into the impact of the subarray structures on the overall system performance.

We propose a criterion to construct the optimal subarrays that maximize a proxy of the system spectral efficiency, i.e., the best partitioning/grouping of the antennas over the RF chains. Using this criterion, we propose a dynamic subarray structure that adapts the subarray structure according to the long-term channel statistics. Finding the optimal subarrays requires an exhaustive search over many antenna partitioning solutions. To lower the complexity, we propose a greedy algorithm that approaches the spectral efficiency of the optimal exhaustive search solution.

The proposed hybrid precoding designs were also evaluated by simulations. Results show that the developed wideband hybrid precoding design approaches the spectral efficiencies of the fully-digital solutions for both fully-connected and fixed-subarray architectures. For the dynamic subarrays, results indicate that their performance outperforms any fixed subarray structure, promoting their potential advantages in wideband mmWave systems.

**Notation:** We use the following notation throughout this paper: $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, $a$ is a scalar, and $\mathcal{A}$ is a set. $|a|$ and $\angle a$ are the magnitude and phase of the complex number $a$. $\|\mathbf{A}\|_F$ is its Frobenius norm, and $\mathbf{A}^T$, $\mathbf{A}^*$, and $\mathbf{A}^{-1}$ are its transpose, Hermitian (conjugate transpose), and inverse, respectively. $[\mathbf{A}]_{1:k}$ denotes the matrix that is composed of the first $k$ columns of the matrix $\mathbf{A}$. $\text{diag}(\mathbf{a})$ is a diagonal matrix with the entries of $\mathbf{a}$ on its diagonal, and $\text{blkdiag}(\mathbf{a}_1, \cdots, \mathbf{a}_k)$ is a block diagonal matrix with $\mathbf{a}_i$’s on its diagonal blocks. $[\mathbf{A}]_{m,n}$ is the $(m, n)$-th element of the matrix $\mathbf{A}$. $\angle \mathbf{A}$ is a matrix with the $(m, n)$-th element equals $e^{j[\mathbf{A}]_{m,n}}$. $\mathbf{I}$ is the identity matrix and $\mathbf{1}_N$ is the $N$-dimensional all-ones vector. $\mathcal{CN}(\mathbf{m}, \mathbf{R})$ is a complex Gaussian random vector with mean $\mathbf{m}$ and covariance $\mathbf{R}$. $\mathbb{E}[\cdot]$ is used to denote expectation.
II. SYSTEM AND CHANNEL MODELS

In this section, we introduce the adopted system and channel models for wideband hybrid precoding.

A. System Model

Consider the system model in Fig. 1. A base station (BS) with $N_{TX}$ antennas and $N_{RF}$ RF chains, $N_{RF} \leq N_{TX}$, communicates with a mobile user that has $N_{RX}$ antennas, via $S$ streams. We assume in this paper that the number of RF chains at the mobile user is equal to the number of antennas, and focus on the hybrid precoding design at the BS. We adopt a hybrid precoding MIMO-OFDM transmission model, similar to [18] with $K$ subcarriers. Let $F_{RF}$ be an $N_{TX} \times N_{RF}$ wideband analog RF precoding matrix, and $F_{BB}[k]$ be an $N_{RF} \times S$ matrix that represents the digital baseband precoding at the $k$-th subcarrier. The transmitted signal can be expressed as

$$x[k] = F_{RF}F_{BB}[k]|s[k], \text{ for } k = 1, \ldots, K,$$

where $s[k]$ is the $S \times 1$ vector of transmitted symbols at subcarrier $k$ with $E[s[k]|s^*[k]] = I_S$. While the analog RF precoding, $F_{RF}$, is performed in the time domain and the same precoding matrix is applied for the entire bandwidth, the digital baseband precoding, $F_{BB}[k]$, is performed in the frequency domain on a per-subcarrier basis. This is a main distinguishing feature of OFDM-based hybrid precoding compared with fully-digital precoding. The precoders, $F_{RF}$ and
$\mathbf{F}_{BB}[k]$, are coupled together through the total power constraints, $\sum_{k=1}^{K} ||\mathbf{F}_{RF}\mathbf{F}_{BB}[k]||_F^2 \leq P_{\text{tot}}$, where $P_{\text{tot}}$ is the total transmit power.

At the receiver, assuming perfect carrier and frequency offset synchronization, the cyclic prefix of length $D$ is first removed from the received signal. The symbols at each subcarrier $k$ are then combined using the $N_{\text{RX}} \times N_{\text{S}}$ digital combining matrix $\mathbf{W}[k]$. Note that no hybrid combining is assumed as the number of RF chains at the receiver equals the number of antennas. Let the $N_{\text{RX}} \times N_{\text{TX}}$ matrix $\mathbf{H}[k]$ denote the channel at subcarrier $k$, the received signal at subcarrier $k$ after processing can be then written as

$$\mathbf{y}[k] = \mathbf{W}^*[k]\mathbf{H}[k]\mathbf{F}_{RF}\mathbf{F}_{BB}[k]\mathbf{s}[k] + \mathbf{W}^*[k]\mathbf{n}[k],$$

(2)

where $\mathbf{n}[k] \sim \mathcal{CN}(0, \sigma_n^2\mathbf{I})$ is the Gaussian noise vector at the receiver.

**B. Channel Model**

We adopt a geometric channel model to incorporate the wideband and limited scattering characteristics of mmWave channels [14], [15], [19]. Consider a mmWave channel with $N_{\text{CH}}$ paths between the BS and mobile user, and let $\rho_p, \tau_p, \phi_{R,p}, \theta_{R,p}, \phi_{T,p}, \theta_{T,p}$ denote the $p$th complex path gain, delay, azimuth angle of departure, elevation angle of departure, azimuth angle of arrival, and elevation angle of arrival, respectively. Let $p(\tau)$ denote a pulse shaping filter for $T_s$-spaced signaling at $\tau$ seconds. The delay-$d$ MIMO channel matrix can be written as [18], [20]

$$\mathbf{H}[d] = \sum_{p=1}^{N_{\text{CH}}} \alpha_p p(dT_s - \tau_p)\mathbf{a}_R(\phi_{R,p}, \theta_{R,p})\mathbf{a}_T^*(\phi_{T,p}, \theta_{T,p}),$$

(3)

where $\mathbf{a}_T(\phi_{T,p}, \theta_{T,p})$ and $\mathbf{a}_R(\phi_{R,p}, \theta_{R,p})$ represent the transmit and receive array response vectors, which depend on the antenna array type. Assuming perfect synchronization, the channel frequency response matrix at each subcarrier $k$ can be expressed as

$$\mathbf{H}[k] = \sum_{p=1}^{N_{\text{CH}}} \alpha_p \omega_{\tau_p}[k]\mathbf{a}_R(\phi_{R,p}, \theta_{R,p})\mathbf{a}_T^*(\phi_{T,p}, \theta_{T,p}),$$

(4)

where $\omega_{\tau_p}[k]$ is defined as

$$\omega_{\tau_p}[k] = \sum_{d=0}^{D-1} p(dT_s - \tau_p)e^{-j2\pi kd}.\quad (5)$$

The channel matrix $\mathbf{H}[k]$ can also be written in a more compact form as

$$\mathbf{H}[k] = \mathbf{A}_R \mathbf{D}[k] \mathbf{A}_T^*,\quad (6)$$
where \( \mathbf{A}_R \) and \( \mathbf{A}_T \) carry the array response vectors of the transmitter and receiver as

\[
\mathbf{A}_R = \begin{bmatrix} \mathbf{a}_R(\phi_{R,1}, \theta_{R,1}) & \mathbf{a}_R(\phi_{R,2}, \theta_{R,2}) & \cdots & \mathbf{a}_R(\phi_{R,N_{CH}}, \theta_{R,N_{CH}}) \end{bmatrix}
\]

\[
\mathbf{A}_T = \begin{bmatrix} \mathbf{a}_T(\phi_{T,1}, \theta_{T,1}) & \mathbf{a}_T(\phi_{T,2}, \theta_{T,2}) & \cdots & \mathbf{a}_T(\phi_{T,N_{CH}}, \theta_{T,N_{CH}}) \end{bmatrix},
\]

and the diagonal matrix \( \mathbf{D}[k] \) equals

\[
\mathbf{D}[k] = \begin{bmatrix} \alpha_1 \omega_{\tau_1}[k] & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \alpha_{N_{CH}} \omega_{\tau_{N_{CH}}}[k] \end{bmatrix}.
\]

In the next section, we formulate the hybrid precoding design problem, before presenting our solutions in the following sections.

### III. Problem Formulation

The objective of this paper is to design the hybrid analog and digital precoders at the BS to maximize the mutual information assuming that the transmit symbol at each subcarrier, \( s[k] \), has a Gaussian distribution. This problem can be formulated as

\[
\left\{ \mathbf{F}_{RF}^*, \{ \mathbf{F}_{BB}^*[k] \} \right\}_{k=1}^{K} = \arg \max_{\mathbf{F}_{RF}, \{ \mathbf{F}_{BB}[k] \}} \sum_{k=1}^{K} \log \det \left( \mathbf{I} + \frac{1}{\sigma_N^2} \mathbf{H}[k] \mathbf{F}_{RF} \mathbf{F}_{BB}[k] \mathbf{F}_{BB}^*[k] \mathbf{F}_{RF}^* \mathbf{H}^*[k] \right)
\]

\[
\text{s.t. } \sum_{k=1}^{K} \| \mathbf{F}_{RF} \mathbf{F}_{BB}[k] \|_F^2 \leq P_{\text{tot}},
\]

where the precoders must satisfy a total power constraint. One difficulty in solving (9) is the coupling between the baseband and RF precoders in the power constraint. Using a change of variable trick, though, and setting \( \mathbf{F}_{BB}[k] = (\mathbf{F}_{RF}^* \mathbf{F}_{RF})^{-\frac{1}{2}} \hat{\mathbf{F}}_{BB}[k] \) with \( \hat{\mathbf{F}}_{BB}[k] \) a dummy variable, the problem in (9) can be equivalently written as [18]

\[
\left\{ \mathbf{F}_{RF}^*, \{ \hat{\mathbf{F}}_{BB}[k] \} \right\}_{k=1}^{K} = \arg \max_{\mathbf{F}_{RF}, \{ \hat{\mathbf{F}}_{BB}[k] \}} \sum_{k=1}^{K} \log \det \left( \mathbf{I} + \frac{1}{\sigma_N^2} \mathbf{H}_{\text{eff}}[k] \hat{\mathbf{F}}_{BB}[k] \hat{\mathbf{F}}_{BB}^*[k] \mathbf{H}_{\text{eff}}^*[k] \right)
\]

\[
\text{s.t. } \sum_{k=1}^{K} \| \hat{\mathbf{F}}_{BB}[k] \|_F^2 \leq P_{\text{tot}},
\]

where \( \mathbf{H}_{\text{eff}} \) is an effective channel matrix defined as

\[
\mathbf{H}_{\text{eff}}[k] = \mathbf{H}[k] \mathbf{F}_{RF} (\mathbf{F}_{RF}^* \mathbf{F}_{RF})^{-\frac{1}{2}}.
\]
If $\mathbf{F}_{RF}$ is given, and assuming perfect channel knowledge at the transmitter, the digital precoders can be found by using a conventional singular value decomposition (SVD) scheme with respect to the effective channel at each subcarrier. Let $\mathbf{H}_{\text{eff}}[k]$ be decomposed by SVD as

$$\mathbf{H}_{\text{eff}}[k] = \mathbf{U}_{\text{eff}}[k] \mathbf{S}_{\text{eff}}[k] \mathbf{V}^*_{\text{eff}}[k],$$

(12)

and let $\mathbf{P}_{\text{eff}}[k]$ be a diagonal matrix whose diagonal elements represent the water-filling power control solution with respect to the effective channel singular values. Then, the optimum solution of $\hat{\mathbf{F}}_{\text{BB}}[k]$ can be represented as

$$\hat{\mathbf{F}}_{\text{BB}}^*[k] = \mathbf{V}_{\text{eff}}[k] \mathbf{P}^{\frac{1}{2}}_{\text{eff}}[k].$$

(13)

Once the optimal $\hat{\mathbf{F}}_{\text{BB}}^*[k]$ is found, the optimal baseband precoder can be calculated as

$$\mathbf{F}_{\text{RF}}^*[k] = (\mathbf{F}_{\text{RF}}^* \mathbf{F}_{RF})^{-\frac{1}{2}} \hat{\mathbf{F}}_{\text{BB}}^*[k] = (\mathbf{F}_{RF}^* \mathbf{F}_{RF})^{-\frac{1}{2}} \mathbf{V}_{\text{eff}}[k] \mathbf{P}^{\frac{1}{2}}_{\text{eff}}[k].$$

(14)

Since the optimal baseband precoding matrices $\mathbf{F}_{\text{BB}}^*[k]$'s depend only on $\mathbf{H}[k]$ and $\mathbf{F}_{RF}$, we can now rewrite the optimization problem in (10) over $\mathbf{F}_{RF}$ only as

$$\mathbf{F}_{\text{RF}}^* = \arg \max_{\mathbf{F}_{RF}} \frac{K}{S} \sum_{k=1}^{K} \sum_{s=1}^{S} \log \left( 1 + \frac{\lambda_s^2(\mathbf{H}[k] \mathbf{F}_{RF}(\mathbf{F}_{RF}^* \mathbf{F}_{RF})^{-\frac{1}{2}}) p_{s,k}}{\sigma_N^2} \right),$$

(15)

where $S = \min(N_{\text{RF}}, N_{\text{RX}})$ is the maximum possible number of streams, $\lambda_s(A)$ is the $s$-th singular value of $A$, and $p_{s,k}$ is the power of the $s$-th stream at the $k$-th subcarrier, which is given by the water-filling power control solution

$$p_{s,k} = \left( \mu - \frac{\sigma_N^2}{\lambda_s^2(\mathbf{H}[k] \mathbf{F}_{RF}(\mathbf{F}_{RF}^* \mathbf{F}_{RF})^{-\frac{1}{2}})} \right)^+, \quad \text{(16)}$$

with $\mu$ satisfying

$$\sum_{k=1}^{K} \sum_{s=1}^{S} \left( \mu - \frac{\sigma_N^2}{\lambda_s^2(\mathbf{H}[k] \mathbf{F}_{RF}(\mathbf{F}_{RF}^* \mathbf{F}_{RF})^{-\frac{1}{2}})} \right)^+ = P_{\text{tot}}. \quad \text{(17)}$$

Note that the original optimization problem in (9) is now equivalent to (15) where we only need to optimize over $\mathbf{F}_{RF}$. This problem, though, is non-convex and hard to solve. Therefore, we relax the optimization and instead maximize the sum of the squared singular values of the effective channels. In Section VII-A, we will evaluate this relaxation and show that it works well for wideband mmWave channels with practical system and channel parameters. Our relaxed objective is to solve

$$\mathbf{F}_{\text{RF}}^* = \arg \max_{\mathbf{F}_{RF}} \frac{K}{S} \sum_{k=1}^{K} \sum_{s=1}^{S} \lambda_s^2(\mathbf{H}[k] \mathbf{F}_{RF}(\mathbf{F}_{RF}^* \mathbf{F}_{RF})^{-\frac{1}{2}}).$$

(18)
Once the optimal RF precoder is found, the water-filling power control is applied with respect to the effective channel singular values associated with $F_{RF}^\star$.

It is worth noting here that we have not put any constraints on the implementation of the RF precoders. Typically, the RF precoding is realized using networks of phase shifters with certain hardware limitations, e.g., only constant-modulus and quantized angles may be allowed. These limitations impose additional constraints on the entries of the RF precoding matrix. This will be addressed later in Section IV after investigating the more relaxed version in (18) with no RF hardware constraints.

IV. WIDEBAND HYBRID PRECODING DESIGN FOR FULLY-CONNECTED ARCHITECTURES

In this section, we consider the system model in Fig. 1 assuming a fully-connected hybrid architecture where each RF chain is connected to all the $N_{TX}$ antennas. In the following proposition, we derive the structure of the optimal RF precoders that solve (18).

\textbf{Proposition 1:} Let $R = \frac{1}{K} \sum_{k=1}^{K} H^*[k]H[k]$ represent a sample covariance matrix of frequency domain channel vectors, with eigenvalue decomposition $R = V_R \Lambda_R V_R^\star$. Let $[V_R]_{1:N_{RF}}$ denote the matrix with the dominant $N_{RF}$ eigenvectors of $R$. The solution to (18) can then be written as

$$ F_{RF}^\star = [V_R]_{1:N_{RF}} A, \tag{19} $$

with an arbitrary $N_{RF} \times N_{RF}$ full rank matrix $A$.

\textbf{Proof: } Let $F_{RF}$ be decomposed by SVD as $F_{RF} = U_{RF} \Lambda_{RF} V_{RF}^\star$. Then, we note that the objective function in (18) can be also written as

$$ \begin{align*}
    &\sum_{k=1}^{K} \sum_{s=1}^{S} \lambda^2_s (H[k]F_{RF}(F_{RF}^\star F_{RF})^{-\frac{1}{2}}) \\
    &= \sum_{k=1}^{K} \|H[k]F_{RF}(F_{RF}^\star F_{RF})^{-\frac{1}{2}}\|^2_F \\
    &= \text{Tr} \left( (F_{RF}^\star F_{RF})^{-\frac{1}{2}}F_{RF}^\star \left( \sum_{k=1}^{K} H^*[k]H[k] \right) F_{RF}(F_{RF}^\star F_{RF})^{-\frac{1}{2}} \right) \\
    &= \text{Tr} \left( F_{RF}(F_{RF}^\star F_{RF})^{-1}F_{RF}^\star \left( \sum_{k=1}^{K} H^*[k]H[k] \right) \right) \\
    &= \text{Tr} \left( KU_{RF}U_{RF}^\star R \right) \\
    &= K\|R^\frac{1}{2}U_{RF}\|^2_F.
\end{align*} \tag{20} $$


Given the eigenvalue decomposition \( R = V_R A_R V_R^* \), the singular vector matrix \( U_{RF}^* \) that maximizes the objective function in (18), or equivalently (25), can now be directly obtained as

\[
U_{RF}^* = [V_R]|_{:N_{RF}} U_A,
\]

(26)

where \( U_A \) is an arbitrary \( N_{RF} \times N_{RF} \) unitary matrix that represents the unitary invariance property of the precoding matrix. Given \( U_{RF}^* \), the optimal RF precoding matrix \( F_{RF} \) that solves (18) can be expressed as

\[
F_{RF}^* = [V_R]|_{:N_{RF}} U_A A_R V_{RF}^*
= [V_R]|_{:N_{RF}} A,
\]

(27)

(28)

where \( A \) is an arbitrary \( N_{RF} \times N_{RF} \) matrix with full rank.

Next, we show that the solution in Proposition 1 achieves the same spectral efficiency as the fully-digital solution to (9) if the number of RF chains is larger than or equal to the number of channel paths, i.e., \( N_{RF} \geq N_{CH} \). First, we rewrite the sample covariance matrix \( R \) as

\[
R = \frac{1}{K} \sum_{k=1}^{K} H^*[k] H[k]
= \frac{1}{K} \sum_{k=1}^{K} (A_T D^*[k] A_{RF}^* A_R D[k] A_T^*)
= A_T \left( \frac{1}{K} \sum_{k=1}^{K} D^*[k] A_{RF}^* A_R D[k] \right) A_T^*.
\]

(30)

(31)

Note that \( A_T \) is an \( N_{TX} \times N_{CH} \) matrix, and \( \sum_{k=1}^{K} D^*[k] A_{RF}^* A_R D[k] \) is an \( N_{CH} \times N_{CH} \) matrix. As a result, the rank of the matrix \( R \) is at most \( \min(N_{TX}, N_{CH}) \). As mmWave systems will employ large antenna arrays [5], [6], and mmWave channels are expected to be sparse [14], [21], the number of channel paths will likely be less than the number of antennas, i.e., \( N_{CH} < N_{TX} \). In this case, the rank of the channel covariance equals the number of paths, i.e., the matrix \( R \) becomes rank-deficient. Based on that, the channel matrix at subcarrier \( k \) can be represented as

\[
H[k] = \hat{H}[k] V_R^*,
\]

(32)

where \( V_R \) is the \( N_{TX} \times N_{CH} \) right singular matrix of \( R \), and \( \hat{H}[k] \) is an \( N_{RX} \times N_{CH} \) matrix. Given that, the fully digital precoding solution that solves the optimization problem in (9) is
given by the SVD solution. Let $H[k] = U[k] \Lambda[k] V^*[k]$ define the SVD of the channel matrix $H[k]$, then the fully-digital optimal precoder equals $V[k]$, which can be written as
\[ V[k] = V_R \tilde{H}^*[k] U[k] \Lambda^{-1}[k]. \] (33)

Assuming that the number of RF chains is at least as large as the number of paths, i.e., $N_{RF} \geq N_{CH}$, then the matrix $V[k]$ in (33) can also be rewritten in terms of the derived baseband and RF precoders in Proposition 1 as
\[ V[k] = F_{RF} F_{BB}[k], \] (34)

with $F_{RF} = V_R A$ and $F_{BB}[k] = A^{-1} \tilde{H}^*[k] U[k] \Lambda^{-1}[k]$. This means that the derived hybrid precoding solution in Proposition 1 represents an optimal solution for (9), and achieves the spectral efficiency of the fully-digital architecture when $N_{RF} \geq N_{CH}$.

To account for the RF constraints, we approximate the unconstrained RF precoder design in (27) by the constrained precoder $\hat{F}_{RF}$ that solves
\[ \hat{F}_{RF} = \arg \min_{X, ||X||_{m,n} = 1} ||X - F_{RF}||_F^2, \] (35)

which is known to provide a good approximation [3], [18]. The solution of (35) is given by $[\hat{F}_{RF}]_{m,n} = e^{j\angle([F_{RF}]_{m,n})}$, where $\angle(\alpha)$ denotes the phase of a complex number $\alpha$. Thanks to the design of the optimal unconstrained RF precoder in (27), which depends on the channel singular vectors, and because these singular vectors take a DFT structure for uniform arrays as $N \to \infty$ [18], [22], [23], this simple solution can be a reasonable substitute for the unknown optimal solution which needs further study. This will be shown by numerical simulations in Section VII.

V. WIDEBAND HYBRID PRECODING DESIGN FOR FIXED SUBARRAY ARCHITECTURES

In this section, we consider the hybrid architecture in Section III but assuming a subarray structure [3], [12]. This means that every RF chain is connected to only a subset of the antennas with $N_{sub} = \frac{N_{TX}}{N_{RF}}$ elements. We assume that $N_{TX}$ is a multiple of $N_{RF}$. Let the antenna indexes be $\{1, \cdots, N_{TX}\}$ and $S_r$ denote the partitioned subset of antenna indexes connected to the $r$-th RF chain such as
\[
S_1 = \{1, \cdots, N_{sub}\} \\
S_2 = \{N_{sub} + 1, \cdots, 2N_{sub}\} \\
\vdots \\
S_{N_{RF}} = \{(N_{RF} - 1)N_{sub} + 1, \cdots, N_{RF}N_{sub}\}.
\] (36)
With this architecture, the analog RF precoding matrix, \( F_{RF} \), has the form of a block diagonal matrix as
\[
F_{RF} = \begin{bmatrix}
\left| f_{RF,S_1} \right| & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \left| f_{RF,S_{N_{RF}}} \right|
\end{bmatrix},
\]
(37)
where \( f_{RF,S_r} \) is an \( N_{sub} \times 1 \) analog beamforming vector associated with the \( r \)-th RF chain. This is a distinct property compared to the fully-connected case whose analog precoding matrix takes the form \( F_{RF} = \begin{bmatrix} f_{RF,1} & f_{RF,2} & \cdots & f_{RF,N_{RF}} \end{bmatrix} \), with \( f_{RF,r} \) an \( N_{TX} \times 1 \) analog beamforming vector associated with the \( r \)-th RF chain. Given this subarray architecture, the overall \( N_{RX} \times N_{TX} \) channel matrix can be expressed using each subarray channel matrix as
\[
H[k] = \begin{bmatrix} H_{S_1}[k] & H_{S_2}[k] & \cdots & H_{S_{N_{RF}}}[k] \end{bmatrix},
\]
(38)
where \( H_{S_r}[k] \) is the \( N_{RX} \times N_{sub} \) channel matrix of the \( r \)-th subarray.

**Proposition 2:** The \( N_{TX} \times N_{RF} \) RF precoder \( F_{RF} \) that solves (18) under the subarray hybrid analog/digital architecture is given by
\[
F_{RF}^* = \text{blkdiag} \left( f_{RF,S_1}^*, \ldots, f_{RF,S_{N_{RF}}}^* \right),
\]
with
\[
f_{RF,S_r}^* = \alpha_r v_{R_{S_r},1}, \quad \text{for} \quad r = 1, \ldots, N_{RF},
\]
(39)
where \( \alpha_r \) is an arbitrary complex value, and \( v_{R_{S_r},1} \) is the largest singular vector of the covariance matrix \( R_{S_r} \), which is associated with the \( r \)-th subarray channel matrix and is defined as
\[
R_{S_r} = \frac{1}{K} \sum_{k=1}^{K} H_{S_r}^*[k] H_{S_r}[k], \quad \text{for} \quad r = 1, \ldots, N_{RF}.
\]
(40)

**Proof:** From (37), \((F_{RF}^* F_{RF})^{-\frac{1}{2}}\) has a form of a diagonal matrix as
\[
(F_{RF}^* F_{RF})^{-\frac{1}{2}} = \begin{bmatrix}
\left| f_{RF,S_1} \right|^{-1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \left| f_{RF,S_{N_{RF}}} \right|^{-1}
\end{bmatrix}.
\]
(41)
This property of \((F_{RF}^* F_{RF})^{-\frac{1}{2}}\) implies that the effective channel for the \( k \)-th subcarrier \( H_{eff}[k] \) in (11) can be written as
\[
H_{eff}[k] = H[k] F_{RF} (F_{RF}^* F_{RF})^{-\frac{1}{2}}
\]
\[
= \begin{bmatrix}
H_{S_1}[k] f_{RF,S_1} & H_{S_2}[k] f_{RF,S_2} & \cdots & H_{S_{N_{RF}}}[k] f_{RF,S_{N_{RF}}}
\end{bmatrix},
\]
(42)
From (40) and (42), the objective function of the optimization problem in (18) can be written as

$$\sum_{k=1}^{K} \sum_{s=1}^{S} \lambda_s^2 \left( H[k] F_{RF}^* (F_{RF} F_{RF})^{-1/2} \right) = \sum_{k=1}^{K} \left| H_{\text{eff}}[k] \right|^2$$

(43)

$$= \sum_{k=1}^{K} \sum_{r=1}^{N_{RF}} \left| H_S[k] f_{RF,S_r} \right|^2 \left| f_{RF,S_r} \right|^2$$

(44)

$$= \sum_{r=1}^{N_{RF}} K f_{RF,S_r}^* R_{S_r} f_{RF,S_r} \left| f_{RF,S_r} \right|^2$$

(45)

where the third equality comes from (40). The maximum value of the objective function in (43) can then be written as

$$\max_{F_{RF}} \sum_{k=1}^{K} \sum_{s=1}^{S} \lambda_s^2 \left( H[k] F_{RF}^* (F_{RF} F_{RF})^{-1/2} \right) = \max_{f_{RF,S_1}, \ldots, f_{RF,S_{N_{RF}}}} \sum_{r=1}^{N_{RF}} K f_{RF,S_r}^* R_{S_r} f_{RF,S_r}$$

(46)

$$= K \sum_{r=1}^{N_{RF}} \lambda_1 (R_{S_r})$$

(47)

where $\lambda_1 (A)$ denotes the largest singular value of a matrix $A$. This maximum value is achieved when the analog beamforming vector for each RF chain $r$ has the structure

$$f_{RF,S_r}^* = \alpha_r v_{R_{S_r}}, \text{ for } r = 1, \ldots, N_{RF}$$

(48)

where $\alpha_r$ is an arbitrary complex value, and $v_{R_{S_r}}$ is the largest singular vector of $R_{S_r}$.

Note that the maximum value of the objective function in (47) is the sum of the largest singular values of $N_{RF}$ submatrices, $R_{S_1}, \ldots, R_{S_{N_{RF}}}$. This is a distinguishing feature from the fully-connected case where the maximum value is the sum of largest $N_{RF}$ singular values of the total matrix, $R$, as

$$\max_{F_{RF}} \sum_{k=1}^{K} \sum_{s=1}^{S} \lambda_s^2 \left( H[k] F_{RF}^* (F_{RF} F_{RF})^{-1/2} \right) = K \sum_{r=1}^{N_{RF}} \lambda_r (R)$$

(49)

While the value of (49) is constant if $R$ is given, the value of (47) depends on the configuration of the submatrices, $R_{S_1}, \ldots, R_{S_{N_{RF}}}$. This motivates a dynamic subarray technique, which will be explained in the next section.

VI. WIDEBAND HYBRID PRECODING DESIGN FOR DYNAMIC SUBARRAY ARCHITECTURES

The subarray hybrid precoding architecture adopted in Section V is the conventional one discussed in prior work [12], [13], where each RF chain is connected to a fixed set of adjacent...
antenna elements. Section V shows that the optimal value of the relaxed mutual information objective function in (46) depends on the sum of the largest singular values of the sample covariance matrices associated to these subarrays. As these largest singular values rely on the selected antennas in each of these subsets, then the question that arises is how much gain can be obtained if these subarrays are dynamically adapted to the long-term channel conditions? Implementing the switch matrix required for the dynamic subarrays shown in Fig. 2(c) is a topic for future work. The objective of this section is to explore the potential gains in terms of the system spectral efficiency.

Now, we define the dynamic subarray problem. We want an algorithm to partition a set of $N_{\text{TX}}$ antennas into $N_{\text{RF}}$ non-empty subsets to maximize the sum of the largest singular values of the submatrices in (47). Note that each antenna should be included only once in one of these subsets, and the union of all these subsets should be the total set of all antenna indexes, $\{1, ..., N_{\text{TX}}\}$. Contrary to the fixed subarray architectures, each subset $S_r$ can have different cardinalities in the dynamic subarray structure. Then, this dynamic subarray partitioning problem to maximize the objective function in (18) can be formulated as

$$\{S_r^*\}_{r=1}^{N_{\text{RF}}} = \arg \max_{S_1, ..., S_{N_{\text{RF}}}} \sum_{r=1}^{N_{\text{RF}}} \lambda_1 (R_{S_r})$$

with

$$s.t. \bigcup_{r=1}^{N_{\text{RF}}} S_r = \{1, \cdots, N_{\text{TX}}\}, S_i \cap S_j = \emptyset \text{ for } i \neq j, \quad |S_r| > 0 \quad \forall r.$$

The problem in (50) is a combinatorial optimization problem for which finding the optimal
solution requires an exhaustive search for all possible cases. The total number of combinations is known as Stirling number of the second kind and is given by
\[
\frac{1}{(N_{\text{RF}})!} \sum_{k=0}^{N_{\text{RF}}} (-1)^{N_{\text{RF}}-k} \binom{N_{\text{RF}}}{k} k^{N_{\text{TX}}},
\] (51)
which is a large number even for a small number of antennas and RF chains. For example, this number becomes \(1.7 \times 10^9\) even for 16 transmit antennas and 4 RF chains. One possible suboptimal solution is to assume that all the subsets have the same size, \(|S_r| = N_{\text{TX}}/N_{\text{RF}}, \forall r\).

Even in this case, though, the total number of combinations is given by
\[
\frac{(N_{\text{TX}})!}{((N_{\text{RF}}^N_{\text{TX}}/N_{\text{RF}}))^{N_{\text{RF}}} (N_{\text{RF}})!},
\]
which is still large, e.g., \(2.6 \times 10^6\) even for 16 transmit antennas and 4 RF chains.

The objective of this section is to develop a low-complexity yet reasonable solution to the problem in (50). First, we note that in many cases, calculating the largest singular values, which is required in (50), does not have a closed form expression and must be calculated numerically, e.g. through an iterative algorithm [24]. Having a closed-form expression of the largest singular value in (50) is important for our subarrays selection problem. To address this challenge, we propose to use a normalized Minkowski \(\ell_1\)-norm [25], which gives a good approximation of the largest singular value as will be discussed in Proposition 3. Given the overall channel covariance matrix \(\mathbf{R}\), the approximate largest singular value of the subset \(S\) is defined as
\[
 \hat{\lambda}_1 (\mathbf{R}_S) \triangleq \frac{1}{|S|} \sum_{i=1}^{|S|} \sum_{j=1}^{|S|} |[\mathbf{R}_S]_{i,j}| = \frac{1}{|S|} \sum_{i \in S} \sum_{j \in S} |[\mathbf{R}]_{i,j}|,
\] (52)
where \(\sum_{i=1}^{|S|} \sum_{j=1}^{|S|} |[\mathbf{R}_S]_{i,j}|\) is known as the Minkowski \(\ell_1\)-norm of the matrix \(\mathbf{R}_S\) [25].

This approximate value has two useful properties. First, this value lies between the existing lower and upper bounds on the exact value of the largest singular value as will be proved shortly in Proposition 3. These lower and upper bounds on the largest singular value of \(\mathbf{R}_S\) (with real eigenvalues) are given by [24]
\[
\lambda_{1,\text{LB}} (\mathbf{R}_S) \leq \lambda_1 (\mathbf{R}_S) \leq \lambda_{1,\text{UB}} (\mathbf{R}_S),
\] (53)
with the lower and upper bounds
\[
\lambda_{1,\text{LB}} (\mathbf{R}_S) = m + \frac{s}{(|S| - 1)^{\frac{1}{2}}},
\] (54)
\[
\lambda_{1,\text{UB}} (\mathbf{R}_S) = m + s(|S| - 1)^{\frac{1}{2}},
\] where
\[
m = \frac{\text{Tr}(\mathbf{R}_S)}{|S|}, \quad s = \left(\frac{\text{Tr}(\mathbf{R}_S^2)}{|S|} - m^2\right)^{\frac{1}{2}}.
\] (55)
In the next proposition, we prove that the approximate largest singular value in (52) also lies between the existing lower and upper bounds in (54).

**Proposition 3:** The approximate value of the largest singular value in (52) has the same lower and upper bound as those of the exact value of the largest singular value if the matrix is Hermitian with identical diagonal elements.

\[ \lambda_{1,\text{LB}} (R_S) \leq \hat{\lambda}_1 (R_S) \leq \lambda_{1,\text{UB}} (R_S). \]  

(56)

**Proof:** See Appendix A.

Note that the channel covariance matrix is a Hermitian matrix and its diagonal elements tend to be identical if all the antennas are located in the same base station because the path loss term is common to all antennas.

The second property of the approximate largest singular value is that this approximate value is a tight lower bound of the exact value in the exponential correlation model case. Even though this correlation model cannot perfectly describe the characteristics of the correlation matrix in the geometric channel model, this can provide an insight to how close the approximate value in (52) is to the exact value, due to its analytical tractability. The spatial channel covariance matrix in the exponential correlation model is

\[ R_S = \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho^* & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ (\rho^*)^{n-1} & (\rho^*)^{n-2} & \cdots & 1 \end{bmatrix}, \]  

(57)

where \( \rho \) is a complex value whose amplitude is less than or equal to 1. The tight lower bound of the largest singular value in this exponential correlation model is known as

\[ \lambda_1 (R_S) \geq \lambda_{1,\text{LB(exp)}} (R_S), \]  

(58)

where

\[ \lambda_{1,\text{LB(exp)}} (R_S) = \frac{1 + |\rho|}{1 - |\rho|} - \frac{2|\rho| (1 - |\rho||S|)}{|S| (1 - |\rho|)^2}. \]  

(59)

In the next proposition, we show that the approximate largest singular value in (52) can be regarded as a tight lower bound of the exact largest singular value in the exponential correlation model case.
Proposition 4: The approximate value of the largest singular value in (52) is the same as the tight lower bound of the exact value of the largest singular value if the matrix is modeled as the exponential correlation matrix.

\[ \hat{\lambda}_1 (R_S) = \lambda_{1, LB(\text{exp})} (R_S). \] (60)

Proof: [Proof] When \( R_S \) is modeled as an exponential correlation model as in (57), the approximate value in (52) can be calculated as

\[ \hat{\lambda}_1 (R_S) = \frac{1}{|S|} \sum_{i=1}^{|S|} \sum_{j=1}^{|S|} |[R_S]_{i,j}| \]
\[ = \frac{1}{|S|} \left( |S| + 2 \sum_{i=1}^{|S|-1} \sum_{j=1}^{|S|} |\rho|^j \right) \]
\[ = \frac{1 + |\rho|}{1 - |\rho|} - \frac{2|\rho| (1 - |\rho||S|)}{|S| (1 - |\rho|)^2}, \] (61)

which is equal to \( \lambda_{1, LB(\text{exp})} (R_S). \)

We propose a practical algorithm using this approximate value of the largest singular value instead of the exact one. At the initial stage, the absolute values in the upper triangular part of \( R \) matrix are sorted in descending order. Then, according to the sorted order, the following process is performed repeatedly. If the selected element at each iteration stage is \( |[R]_{i,j}| \), then the algorithm checks whether \( i \)-th antenna and \( j \)-th antenna are in the same subset or not. If they are in different subsets, the algorithm tries relocating one antenna to the subset that the other antenna belongs to, and calculates the metric, which is defined as the sum of the proposed approximate largest singular values of submatrices. Note that only at most two subsets can be changed at each stage while other subsets remain unchanged. Therefore, the singular values of other submatrices need not be recalculated, and thus the metric at each stage can be simplified as the sum of the two singular values. If the newly calculated metric is larger than the current metric, then the algorithm decides to relocate the antenna, and otherwise decides to maintain the current status. The pseudo code of the details in the proposed algorithm is shown in Algorithm 1. In Algorithm 1, the function \( f_{R,N_{RF}}(\cdot) \) is defined as

\[ f_{R,N_{RF}} (S, n_{\text{sel}}, r) \triangleq \begin{cases} 
0, & \text{if } |S| = 0 \text{ or } \{ n_{\text{sel}} = N_{RF} \text{ and } r = 0 \} \\
\frac{1}{|S|} \sum_{i \in S} \sum_{j \in S} |[R]_{i,j}|, & \text{otherwise}
\end{cases}, \] (62)

which indicates the approximate singular value of the covariance matrix of the antenna subset.
Algorithm 1 Dynamic subarray partitioning

Input: $R$, $N_{RF}$, $N_{TX}$

$S_0 = \{1, \ldots, N_{RF}\}$, $n_{sel} = 0$

Sort $[R]_{i,j}$ for $1 \leq i < j \leq N_{TX}$ in descending order

$\left( |[R]_{i_1,j_1}| \geq \cdots \geq |[R]_{i_K,j_K}|, \ K = \frac{N_{TX}(N_{TX}-1)}{2}, \ 1 \leq i_k < j_k \leq N_{TX} \right)$

for $k = 1 : K$ do

if $i_k, j_k \in S_0$ then

if $n_{sel} < N_{RF}$ then

$n_{sel} \leftarrow n_{sel} + 1$, $S_{n_{sel}} \leftarrow \{i_k, j_k\}$, $S_0 \leftarrow S_0 \setminus \{i_k, j_k\}$

else

$\hat{r} = \arg \max_{r \in \{1, \ldots, N_{RF}\}} \left( f_{R,N_{RF}}(S_r \cup \{i_k, j_k\}, n_{sel}, r) - f_{R,N_{RF}}(S_r, n_{sel}, r) \right)$

$S_{\hat{r}} \leftarrow S_{\hat{r}} \cup \{i_k, j_k\}$, $S_0 \leftarrow S_0 \setminus \{i_k, j_k\}$

else if $i_k \in S_m, j_k \in S_l$ for some $m, l \in \{0, 1, \ldots, n_{sel}\}$ and $m \neq l$ then

$\mu_{\text{current}} = f_{R,N_{RF}}(S_m, n_{sel}, m) + f_{R,N_{RF}}(S_l, n_{sel}, l)$

$\mu_{\text{new},j} = f_{R,N_{RF}}(S_m \cup \{j_k\}, n_{sel}, m) + f_{R,N_{RF}}(S_l \setminus \{j_k\}, n_{sel}, l)$

$\mu_{\text{new},i} = f_{R,N_{RF}}(S_m \setminus \{i_k\}, n_{sel}, m) + f_{R,N_{RF}}(S_l \cup \{i_k\}, n_{sel}, l)$

if $\mu_{\text{new},j} > \mu_{\text{new},i}$, $\mu_{\text{new},j} > \mu_{\text{current}}$, and $m \neq 0$ then

$S_m \leftarrow S_m \cup \{j_k\}$, $S_l \leftarrow S_l \setminus \{j_k\}$

else if $\mu_{\text{new},i} > \mu_{\text{new},j}$, $\mu_{\text{new},i} > \mu_{\text{current}}$, and $l \neq 0$ then

$S_m \leftarrow S_m \setminus \{i_k\}$, $S_l \leftarrow S_l \cup \{i_k\}$

Output: $S_1, \ldots, S_{N_{RF}}$

VII. SIMULATION RESULTS

In this section, we first evaluate the performance of the proposed wideband hybrid precoding design in a mmWave frequency selective channel, and then present simulation results to demonstrate the performance of the proposed dynamic subarray algorithm with hybrid architectures.

In the simulations, we consider the channel model in Section II. The channel is modeled as a clustered channel where each cluster is composed of multiple subrays. The distributions of the paths’ delay and azimuth/elevation angles are similar to that in the 3GPP 3D-MIMO channel model [27] and WINNER II SCM channel model [28]. Considering multiple rays per cluster,
we can rewrite the channel model in (4) as

\[ H[k] = \sum_{c=1}^{N_{\text{cluster}}} \sum_{r=1}^{N_{\text{subray}}} \alpha_{c,r} \omega_{c,r} [k] a_R(\phi_{R,c,r}, \theta_{R,c,r}) a_T^*(\phi_{T,c,r}, \theta_{T,c,r}), \]  

(63)

Unless otherwise mentioned, the adopted channel has \( N_{\text{cluster}} = 8 \) clusters whose center azimuth angles of arrival and departure are uniformly distributed in \([-180^\circ, 180^\circ]\), and the center elevation angles of arrival and departure are uniformly distributed in \([-90^\circ, 90^\circ]\) when UPA is used in the simulation. Each cluster is composed of \( N_{\text{subray}} = 10 \) subrays whose azimuth and elevation angles are assumed to be Laplacian distributed with angular spread of 5\(^\circ\) [28]. Both ULA and UPA types are simulated, and the antenna spacing between antennas is 0.5\(\lambda\), where \(\lambda\) is the signal wavelength. A raised-cosine filter with a roll-off factor one is adopted for the pulse shaping filter. The number of subcarriers \( K \) is 4096, and the cyclic prefix length \( D \) is assumed to be \( K/4 \) as in IEEE 802.11ad. All subrays within a cluster is assumed to have an identical delay such that \( \tau_{c,1} = \cdots = \tau_{c,N_{\text{subray}}} = \tau_c \). The cluster delay \( \tau_c \) normalized to \( T_s \) is assumed to have a discrete uniform distribution in the cyclic prefix duration, \([0, D]\). The water-filling power control policy is used for all test cases.

A. Evaluating the Relaxation of the Optimum Criterion

We used the relaxed optimum criterion in (18) instead of the exact optimum criterion (15). Fig. 3 shows the exact value in the original problem

\[
\frac{1}{KS} \sum_{k=1}^{K} \sum_{s=1}^{S} \log \left( 1 + \frac{\lambda^2_s (H[k])}{\sigma_N^2} \right),
\]  

(64)

and the approximate value in the relaxed criterion

\[
\log \left( 1 + \frac{1}{KS} \sum_{k=1}^{K} \sum_{s=1}^{S} \frac{\lambda^2_s (H[k])}{\sigma_N^2} \right),
\]  

(65)

which is Jensen’s upper bound of (64). It is well known that this bound is tight only at low SNR region. The bound, however, can be also tight even at high SNR if the number of transmit antennas is larger than the number of receive antennas. Fig. 3(a) shows the bound and the exact value in IID Rayleigh fading channel according to the number of transmit antennas when the number of receive antennas is two. It is shown that the gap becomes smaller as the ratio of the number of transmit antennas to the number of receive antennas becomes larger. The gap is negligible when the number of transmit antennas is more than 16 even at high SNR. Fig. 3(b) shows the results of the sparse mmWave channel case in (63), where \( N_{\text{cluster}} = 8, N_{\text{subray}} = 10, \)
Fig. 3. Comparison between the exact value given in (64) and the approximate value given in (65), which is Jensen’s upper bound of (64). The number of receive antennas is fixed at two. IID Rayleigh channel model is assumed in (a), and the mmWave channel model is used in (b).

and ULA type antennas are used at BS. These results indicate that the gap between the bound and the exact value is $\sim 1$ bps/Hz when more than 16 antennas are deployed, which means that the relaxed optimization problem in (18) is a reasonable approximation of the original problem.
in (15) for large MIMO mmWave systems.

B. Wideband Hybrid Precoding over Frequency Selective Channels

In Fig. 4(a) and Fig. 4(b), we evaluate the performance of the proposed wideband hybrid precoding design for SU-MIMO over a frequency selective channel in a fully-connected structure. The figures show the average mutual information per subcarrier according to SNR when the wideband hybrid precoding in Section IV are used in the case of 16 transmit antennas and 4 receive antennas. The number of RF chains at the receiver is 4, and fully-digitalized baseband combining is used. The number of RF chains at the transmitter is 1, 2, 4, or 8. In Fig. 4(a), the channel per subcarrier and per antenna is modeled as IID Rayleigh channel, which is an extreme case of an ideal rich scattering environment. The results show that there is a substantial loss from the fully-digitalized baseband precoding case even when eight RF chains are used in the wideband hybrid precoding. This, however, is not the case when the sparse mmWave channel is considered. Fig. 4(b) shows the wideband hybrid precoding performance when the mmWave channel model in (63) is adopted, with $N_{\text{cluster}} = 8$ and $N_{\text{subray}} = 10$. If each cluster has only one ray and the number of clusters is less than or equal to the number of RF chains, the performance of the proposed wideband hybrid precoding is the same as that of the fully-digital precoding as discussed in Section IV. Even when the channel clusters have multiple subrays with angle spread $5^\circ$, Fig. 4(b) shows that the performance gap between the proposed hybrid precoding and the fully-digital solution is negligible when eight RF chains.

C. Wideband Hybrid Precoding with Dynamic Subarray Structures

In this subsection, we evaluate the performance of the proposed algorithm in the subarray structure. The proposed dynamic subarray technique is compared with several fixed subarray types as well as the fully-connected hybrid precoding and the fully-digitalized baseband precoding. For the dynamic subarray architecture, the proposed greedy algorithm is compared with the optimal exhaustive search algorithm. For comparison, we also simulate another simple technique of dynamic subarrays that selects the best subarray architecture among a predefined fixed subarray types. The simulations are conducted in various channel environments to establish the dependence of the dynamic subarray gain on channel parameters. In addition to evaluating the dynamic subarrays, we also establish which is the best fixed subarray structure and which channel parameters affect the decision of the best structure.
Fig. 4. Comparison between the fully-digitalized baseband precoding and the hybrid precoding with 1, 2, 4, or 8 RF chains in a fully-connected structure. The transmitter uses 16 antennas (ULA), and the receiver uses 4 antennas (ULA). In (a), the channel at each subcarrier is modeled as IID Rayleigh channel. The mmWave channel model used in (b) is the same as in Fig. 3(b).

**Gain over fixed subarrays in ULA systems:** Fig. 5 shows the comparison of various precoding techniques when the base station has 9 antennas (ULA) and 3 RF chains and the mobile station has 2 antennas (ULA) and 2 RF chains. For comparison, two fixed subarray types...
(a) Fixed subarray types for simulations: 9 antennas (1x9 ULA) and 3 RF chains.

![Fixed subarray types](image)

(b) Spectral efficiency vs. SNR.

![Spectral efficiency vs. SNR](image)

Fig. 5. Comparison among various hybrid architectures when 9 antennas (ULA) and 3 RF chains are equipped at BS and 2 antennas (ULA) and 2 RF chains are equipped at MS.

are used in the simulation as described in Fig. 5(a): an adjacent type and an interlaced type. Fig. 5(b) shows that, in the dynamic subarray architecture, the performance of the proposed algorithm is close to the optimal exhaustive search case even with a much lower complexity. The results also indicate that the adjacent type is better than the interlaced one among two fixed subarray structures. This is because the largest singular value of each adjacent type is larger than that of a interlaced type when the channel is correlated.

**Gain over fixed subarrays in UPA systems:** Fig. 6 shows a simulation result when 64 antennas (8x8 UPA) and 4 RF chains are used at the base station and 4 antennas (2x2 UPA) and 4 RF chains are used at the mobile station. Nine fixed subarray types are used in the simulation as shown in Fig. 6(a). The optimal exhaustive search algorithm for the dynamic subarrays was not simulated because its computational complexity is too high in this case. Instead, a simple dynamic subarray algorithm that selects the best subarray structure in a predefined set, which consists of the nine fixed subarray types in Fig. 6(a), was simulated for comparison. Even though
Fig. 6. Comparison among various hybrid architectures when 64 antennas (8x8 UPA) and 4 RF chains are equipped at BS and 4 antennas (2x2 UPA) and 4 RF chains are equipped at MS.

The dynamic subarray technique with a predefined set outperforms any fixed subarray types, this naive dynamic subarray technique is considerably outperformed by the proposed algorithm. This
is because the simple dynamic algorithm selects the best subarray type among nine fixed subarray types while the proposed algorithm can decide the best subarray type among all possible types, whose total number is $1.4 \times 10^{37}$. Compared to the results in Fig. 5, we can see that the gain of the proposed dynamic algorithm becomes higher as the number of antennas and RF chains gets larger. Apart from the dynamic subarrays, the figures shows the information about the best structure when a fixed subarray structure is applied. As can be seen in Fig. 6(b), the squared type is the best structure among the fixed subarray structures, and the vertical type is the second. This trend is consistent with the results in Fig. 6(c) that shows the selection ratio of fixed subarray structures when a simple dynamic algorithm with a predefined set is used.

**Impact of channel parameters:** The best fixed subarray structure as well as the dynamic subarray gain depends the channel environment. In particular, the distributions of azimuth and elevation angles of channel paths play an important role, as the angle distributions affect the largest singular value of each subarray. The azimuth angles and elevation angels can be confined within some range in some cell deployment scenarios. For example, the range of incoming azimuth angles can be restricted in a 3-sectorized cell scenario where sector antennas with directional antenna gain are equipped. In addition, the range of azimuth angles can be different from that of elevation angles. For example, many outdoor scenarios are usually assumed to have a smaller range of elevation angles than that of azimuth angles [28].

Fig. 7 shows the influence of the azimuth and elevation angle range on the objective function in (50), which is the sum of the dominant singular values of each subarray. In Fig. 7(a) the center azimuth angles per cluster are assumed to be uniformly distributed within $[-\phi_{\text{max}}, \phi_{\text{max}}]$, so the maximum azimuth difference from the antenna boresight angle is limited to $\phi_{\text{max}}$. The center elevation angles are assumed to be uniformly distributed within $[-90^\circ, 90^\circ]$, which means that there is no restriction on the elevation angle range. The figure shows that the gain of the dynamic subarray technique increases as the range of angles becomes wider. The figure also demonstrates that the best fixed subarray structure varies according to the angle range. The horizontal fixed type structure outperforms other fixed types when the azimuth angles are confined within a small range. The main reason is due to the difference in the range of angles. If the range in the azimuth angles is narrower than the range in the elevation angles, the largest singular values of the covariance channel matrix of each horizontal row is larger than that of each vertical column, and thus the horizontal fixed type structure has a higher value of the objective function in (50) than the vertical fixed type and others. The squared fixed type, however, becomes the
Fig. 7. The objective function in (50) vs. range of angles. The BS uses 81 antennas (9x9 UPA) and 9 RF chains, and the MS uses 4 antenna (2x2 UPA) and 4 RF chains. The angle spread within a cluster is 5°, and SNR is 10dB. In (a), The center azimuth angles per cluster are uniformly distributed within $[-\phi_{\text{max}}, \phi_{\text{max}}]$, and the center elevation angles per cluster are uniformly distributed within $[-90^\circ, 90^\circ]$. In (b), The center azimuth angles per cluster are uniformly distributed within $[-180^\circ, 180^\circ]$, and the center elevation angles per cluster are uniformly distributed within $[-\theta_{\text{max}}, \theta_{\text{max}}]$. 

best among fixed subarray types as the range of azimuth angles becomes bigger. This is because the horizontal domain and the vertical domain have a similar level of correlation, which enables
Fig. 8. Spectral efficiency vs. number of transmit antennas. The number of RF chains at BS is fixed at four. The MS uses 4 antenna (2x2 UPA) and 4 RF chains. The center azimuth angles per cluster are uniformly distributed in $[-180^\circ, 180^\circ]$, and the center elevation angels per cluster are uniformly distributed in $[-90^\circ, 90^\circ]$, the angles per subray within a cluster are Laplacian distributed with $5^\circ$, and SNR is 10dB. The dashed curves indicate the case of RF precoding with only phase shifters under the constant modulus constraint while the solid ones represent the case of RF precoding without the constraint.

the squared fixed type subarray structure to have the largest singular value due to the smallest distances between antennas. A similar phenomenon occurs when the elevation angles have a limited range. Fig. 7(b) shows that the vertical fixed type outperforms other fixed types when the range of elevation angles is small and the squared fixed type is the best at larger ranges as Fig. 7(a).

**Performance with different antenna array sizes:** The gain of proposed dynamic subarray structure also varies with the number of antennas as can be seen in Fig. 8. In the figure, the MS has 4 antennas (2x2 UPA) and 4 RF chains, and the BS has 16, 64, 144, or 256 antennas (4x4, 8x8, 12x12, or 16x16 UPA). The number of RF chains at BS is assumed to be fixed at four. The figure also shows the impact of the phase shifter constraint. The solid curves represent the unconstrained RF precoding, and the dashed curves indicate the constrained RF precoding with phase shifters in the analog RF precoding. The figure shows that the loss from using the phase shifters in the analog domain is not substantial. Apart from the phase shifter constraint issue, Fig. 8 also shows that the gain of the proposed dynamic structure becomes higher as the
number of transmit antennas becomes larger.

VIII. CONCLUSIONS

In this paper, we developed hybrid analog/digital precoding design for wideband mmWave MIMO-OFDM systems with frequency selectivity. Considering a relaxation of the mutual information maximization problem, we derived a near-optimal closed-form solution for both fully-connected and partially-connected hybrid architectures. Simulation results showed that the spectral efficiency of the proposed wideband hybrid precoding designs approaches that obtained with fully-digital precoding. Inspired by the developed closed-form solution, we explored the potential spectral efficiency gain if the antenna subarrays can be adaptively adjusted according to the large channel statistics. For that, we first developed a criterion for constructing these subarrays, and used to design an antenna partitioning algorithm. One insight, drawn from the developed criterion, is that forming each subarray with more correlated antenna elements normally leads to an efficient subarray structure. Simulation results showed that the achievable spectral efficiency by dynamic subarrays outperforms that of fixed subarray architectures. For future work, it would be interesting to evaluate the trade-off between the achieved spectral efficiency and the consumed energy of the dynamic subarray structure, and compare it with the fully-connected and the fixed-subarray architectures.

APPENDIX A

Proof: [Proof of Proposition 3] Let the diagonal elements of \( \mathbf{R}_S \) be normalized to one and \(|S| = n\). Then, (55) becomes

\[
m = 1, \quad s = \left( \frac{1}{n} ||\mathbf{R}_S||_F^2 - 1 \right)^{\frac{1}{2}} = \left( \frac{2}{n} \sum_{i=1}^{n} \sum_{j>i} ||\mathbf{R}_S||_{i,j}^2 \right)^{\frac{1}{2}}.
\]

Define \( \gamma_k \) for \( k = 1, \ldots, \frac{n(n-1)}{2} \) to be

\[
\gamma_1 = ||\mathbf{R}_S||_{1,2}, \quad \gamma_2 = ||\mathbf{R}_S||_{1,3}, \ldots, \quad \gamma_{n-1} = ||\mathbf{R}_S||_{1,n},
\]

\[
\gamma_n = ||\mathbf{R}_S||_{2,3}, \ldots, \quad \gamma_{2n-3} = ||\mathbf{R}_S||_{2,n},
\]

\[
\gamma_{2n-2} = ||\mathbf{R}_S||_{3,4}, \ldots
\]

\[
\gamma_{n(n-1)} = ||\mathbf{R}_S||_{n-1,n}.
\]
Then, the lower and upper bounds in \((54)\) can be rewritten as

\[
\lambda_{1,\text{LB}}(R_S) = 1 + \left( \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} |(R_S)_{i,j}|^2 \right)^{\frac{1}{2}} = 1 + \left( \frac{2}{n(n-1)} \sum_{k=1}^{n(n-1)/2} \gamma_k^2 \right)^{\frac{1}{2}},
\]

\[
\lambda_{1,\text{UB}}(R_S) = 1 + \left( \frac{2(n-1)}{n} \sum_{i=1}^{n} \sum_{j>i}^{n} |(R_S)_{i,j}|^2 \right)^{\frac{1}{2}} = 1 + \left( \frac{2(n-1)}{n} \sum_{k=1}^{n(n-1)/2} \gamma_k^2 \right)^{\frac{1}{2}}.
\]

The approximate value of the largest singular value in \((52)\) can be represented as

\[
\hat{\lambda}_1(R_S) = 1 + \frac{2}{n} \sum_{i=1}^{n} \sum_{j>i}^{n} |(R_S)_{i,j}| = 1 + \frac{2}{n} \sum_{k=1}^{n(n-1)/2} \gamma_k
\]

The relation between the lower bound of the exact value and the approximate value is

\[
\frac{\lambda_{1,\text{LB}}(R_S) - 1}{\lambda_1(R_S) - 1} = \left( \frac{n}{2(n-1)} \right) \cdot \frac{\sum_{k=1}^{n(n-1)/2} \gamma_k^2}{\left( \sum_{k=1}^{n(n-1)/2} \gamma_k \right)^2} \leq 1 \text{ for } n \geq 2
\]

because

\[
\frac{n}{2(n-1)} \leq 1 \text{ for } n \geq 2
\]

and

\[
\left( \sum_{k=1}^{n(n-1)/2} \gamma_k \right)^2 = \sum_{k=1}^{n(n-1)/2} \gamma_k^2 + 2 \sum_{k=1}^{n(n-1)/2} \gamma_k \gamma_m \geq \sum_{k=1}^{n(n-1)/2} \gamma_k^2,
\]

where the inequality stems from the fact that \(\gamma_k \geq 0, \forall k\) by definition in \((67)\). From \((70)\), we can conclude that

\[
\hat{\lambda}_1(R_S) \geq \lambda_{1,\text{LB}}(R_S) \text{ for } n \geq 2.
\]

Now, consider the relationship between the upper bound of the exact and approximate singular values. This ratio between the two values can be written as

\[
\frac{\lambda_{1,\text{UB}}(R_S) - 1}{\hat{\lambda}_1(R_S) - 1} = \left( \frac{n(n-1)}{2} \cdot \frac{\sum_{k=1}^{n(n-1)/2} \gamma_k^2}{\left( \sum_{k=1}^{n(n-1)/2} \gamma_k \right)^2} \right)^{\frac{1}{2}} \geq 1 \text{ for } n \geq 2
\]
because

\[
\left( \sum_{k=1}^{n(n-1)/2} \gamma_k \right)^2 \leq \left( \sum_{k=1}^{n(n-1)/2} 1 \cdot \gamma_k \right)^2 \leq \left( \sum_{k=1}^{n(n-1)/2} 1^2 \right) \left( \sum_{k=1}^{n(n-1)/2} \gamma_k^2 \right) = \frac{n(n - 1)}{2} \left( \sum_{k=1}^{n(n-1)/2} \gamma_k^2 \right)
\]

which results from Cauchy-Schwarz inequality. Therefore, we get

\[
\hat{\lambda}_1 (R_S) \leq \lambda_{1,UB} (R_S) \quad \text{for} \quad n \geq 2,
\]

which completes the proof. \[\square\]

REFERENCES

[1] X. Zhang, A. Molisch, and S. 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.

[2] V. Venkateswaran and A. van der Veen, “Analog beamforming in MIMO communications with phase shift networks and online channel estimation,” *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[3] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[4] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. Heath, “MIMO precoding and combining solutions for millimeter-wave systems,” *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[5] W. Roh, J. 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, February 2014.

[6] R. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. Sayeed, “An overview of signal processing techniques for millimeter wave MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, April 2016.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, “Hybrid precoding for millimeter wave cellular systems with partial channel knowledge,” in *Proc. of Information Theory and Applications Workshop (ITA)*, Feb 2013, pp. 1–5.

[8] W. Ni and X. Dong, “Hybrid block diagonalization for massive multiuser MIMO systems,” *IEEE Transactions on Communications*, vol. 64, no. 1, pp. 201–211, Jan. 2016.

[9] X. Yu, J. Shen, J. Zhang, and K. Letaief, “Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, April 2016.

[10] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, and R. Heath, “Dictionary-free hybrid precoders and combiners for mmwave MIMO systems,” in *Proc. of the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015.

[11] C. Chen, “An iterative hybrid transceiver design algorithm for millimeter wave MIMO systems,” *IEEE Wireless Communications Letters*, vol. 4, no. 3, pp. 285–288, June 2015.

[12] S. Han, C. 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.

[13] O. El Ayach, R. Heath, S. Rajagopal, and Z. Pi, “Multimode precoding in millimeter wave MIMO transmitters with multiple antenna sub-arrays,” in *Proc. of the IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 3476–3480.

[14] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!” *IEEE Access*, vol. 1, pp. 335–349, May 2013.
[15] M. Samimi and T. Rappaport, “Ultra-wideband statistical channel model for non line of sight millimeter-wave urban channels,” in *Proc. of the IEEE Global Communications Conference (GLOBECOM)*, Dec 2014, pp. 3483–3489.

[16] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.

[17] C. Kim, T. Kim, and J. Seol, “Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems,” in *Proc. of IEEE Globecom Workshops (GC Wkshps)*, Atlanta, GA, Dec. 2013, pp. 61–65.

[18] A. Alkhateeb and R. 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] T. Rappaport, Y. Qiao, J. Tamir, J. Murdock, and E. Ben-Dor, “Cellular broadband millimeter wave propagation and angle of arrival for adaptive beam steering systems,” in *Radio and Wireless Symposium*, Santa Clara, CA, Jan. 2012, pp. 151–154.

[20] P. Schniter and A. Sayeed, “Channel estimation and precoder design for millimeter-wave communications: The sparse way,” in *Proc. of Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 273–277.

[21] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, “Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications,” *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[22] A. Adhikary, J. Nam, J. Ahn, and G. Caire, “Joint spatial division and multiplexing: The large-scale array regime,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[23] O. El Ayach, R. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, “The capacity optimality of beam steering in large millimeter wave MIMO systems,” in *Proc. of IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Cesme, Turkey, June 2012, pp. 100–104.

[24] H. Wolkowicz and G. Styan, “Bounds for eigenvalues using traces,” *Linear Algebra and Its Applications*, no. 29, pp. 471–506, 1980.

[25] H. Lütkepohl, *Handbook of Matrices*. Wiley, 1996.

[26] J. Choi and D. Love, “Bounds on eigenvalues of a spatial correlation matrix,” *IEEE Communications Letters*, vol. 18, no. 8, pp. 1391–1394, Aug 2014.

[27] 3GPP TR 36.873 V12.2.0, “Study on 3D channel model for LTE (Release 12),” June 2015.

[28] P. Kyosti et al. IST-4-027756 WINNER II D1.1.2 V1.2 WINNER II Channel Models. [Online]. Available: www.ist-winner.org