Massive MIMO System Partitioning for Efficient Hybrid Beamformer Optimization

Saidur R. Pavel [1], Yimin D. Zhang [1], and Batu K. Chalise [2]
[1] Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA
[2] Department of Electrical and Computer Engineering, New York Institute of Technology, NY 11568, USA

***Abstract*—Hybrid analog-digital beamforming is an effective approach for practical implementations of a massive multiple-input multiple-output (MIMO) system by reducing the number of radio frequency (RF) chains. Fully connected hybrid beamforming (F-HBF), where each RF chain is connected to each antenna, can lower hardware complexity, power consumption, and cost compared to digital beamforming. Subarray-based hybrid beamforming (S-HBF), where a specific group of RF chains is allocated to a particular subarray, can further reduce hardware requirements. The antenna array is divided into subarrays using effective partitioning so that the optimization of analog beamforming can be shared across multiple subarrays, substantially reducing computational complexity.**

Keywords: Hybrid beamforming, fully connected hybrid beamforming, subarray-based hybrid beamforming, massive MIMO, subarray partitioning, mutual information.

I. Introduction

Massive multiple-input multiple-output (MIMO) is regarded as one of the emerging technologies in the next-generation wireless communication systems by enhancing energy efficiency, system capacity, and quality of service [1–5]. Particularly, in millimeter wave (mmWave) communications, massive MIMO systems can be exploited to form highly directional beams to address the high propagation attenuation problem of mmWave channels [6–9]. Deploying a massive MIMO system, however, poses challenges in terms of hardware implementation, cost, power consumption, and algorithm design. Full digital beamforming requires one dedicated radio frequency (RF) chain per antenna. Each RF chain generally consists of hardware components including signal mixers, analog-to-digital converters (ADCs), and power amplifiers. The requirement of a high number of RF components prohibitively increases hardware cost, power consumption, and the computational complexity of the beamforming algorithm, thus making the system infeasible.

To address such challenges, hybrid analog-digital beamforming techniques [10–18] are popularly employed as a cost-effective solution. Hybrid beamforming significantly reduces the number of RF chains by dividing the entire beamforming process into two parts, namely, high-dimensional analog beamforming and low-dimensional baseband digital beamforming, where the former is implemented using phase shifters. The implementation of hybrid beamforming can vary based on how the RF chains are connected to the antenna elements [15, 19, 20]. In a traditional full hybrid beamforming (F-HBF) system, the analog beamformer connects all antennas with all RF chains. To further reduce the complexity, several hybrid beamforming structures with flexible mapping strategies have been developed to balance hardware complexity and spectral efficiency. One of the effective strategies is subarray-based hybrid beamforming (S-HBF) in which the antennas in a subarray are only connected to the corresponding subset of RF chains [20]. Different subarray partitioning approaches, such as the localized, interleaved, and nested partitioning [19, 21], can be used. In this paper, localized partitioning is considered.

Optimizing the analog beamformer in a hybrid beamforming scheme is a challenging task. While many existing methods assume the availability of full-dimensional analog signals, signal direction of arrivals (DOAs), and/or channel state information when optimizing the analog beamformer, such information is generally unavailable in practice as we only have access to the digital version of the analog beamformer output. Therefore, the analog beamforming matrix must be optimized blindly without the knowledge of the full-dimensional analog signals. To achieve this, an information-theoretic approach is developed in [10, 16], where the analog beamformer, also referred to as the compressive measurement matrix, is optimized by maximizing the mutual information (MI) of the analog beamformer output and signal DOAs. In the absence of prior knowledge about the distribution of the DOAs, [22] develops an iterative approach that initially considers a uniform prior of DOAs and then updates this prior iteratively by estimating the normalized spatial spectrum. In [17, 18, 23, 24], deep neural networks are utilized to optimize the compressive measurement matrix. [25] maximizes the MI criterion by further taking array imperfections into account. However, all of these methods only considers the F-HBF structure.

By exploiting identical subarray partitioning, we optimize the analog beamformer in one of the subarrays and use this knowledge to derive the full beamforming matrix. The optimization of analog beamformer in each subarray is modified from an information-theoretic iterative approach developed in [16, 22] in which the analog beamformer is optimized by maximizing the MI between the analog beamformer output and the signal DOAs. Without assuming prior knowledge of the signal arrivals, it starts with a uniform prior, and the normalized power spectrum density estimate at an iteration is then utilized as the prior DOA distribution for the next iteration.

Notations: We use bold lower-case letters to represent vectors and bold upper-case letters to represent matrices. Specifically, $\boldsymbol{I}$ denotes the identity matrix of the appropriate dimension. The symbols $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ respectively represent the transpose and Hermitian operations of a matrix or vector. In addition, $\mathrm{vec}(\cdot)$ vectorizes a matrix and $\mathrm{Diag}(\cdot)$ forms a matrix by placing a vector in its diagonal elements. $\|\cdot\|_{\mathrm{F}}$ defines the Frobenius norm, while $|\cdot|$ represents the element-wise absolute value of a matrix. The symbols $\otimes$ and $\odot$ denote the Kronecker product and Khatri-Rao product, respectively.

II. Signal Model

Consider Q uncorrelated sources impinging on a massive MIMO system consisting of N antennas arranged in uniform linear fashion

The work of S. R. Pavel and Y. D. Zhang was supported in part by the National Science Foundation (NSF) under grant No. ECCS-2236023.

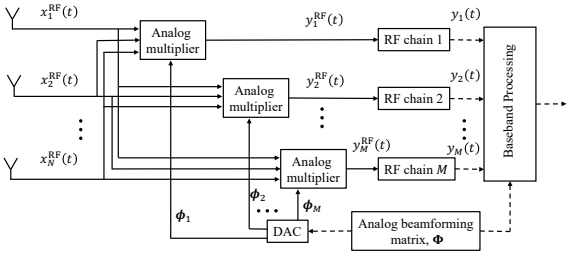


Fig. 1: Fully connected hybrid beamforming structure.

from directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_Q]^{\mathrm{T}}$. The analog RF signal received by the antenna array at time t can be expressed as

$$\begin{aligned} \boldsymbol{x}^{\mathrm{RF}}(t) &= \sum_{q=1}^{Q} \boldsymbol{a}(\theta_q) s_q(t) e^{j\omega_c t} + \boldsymbol{n}^{\mathrm{RF}}(t) \\ &= \boldsymbol{A}(\theta)\boldsymbol{s}(t) e^{j\omega_c t} + \boldsymbol{n}^{\mathrm{RF}}(t), \end{aligned} \tag{1}$$

where $\boldsymbol{A}(\theta) = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \cdots, \boldsymbol{a}(\theta_Q)]^{\mathrm{T}} \in \mathbb{C}^{N\times Q}$ represents the array manifold matrix with $\boldsymbol{a}(\theta_q) = [1, e^{-j\frac{2\pi}{\lambda}d\sin\theta_q}, \cdots, e^{-j\frac{2\pi}{\lambda}(N-1)d\sin\theta_q}]^{\mathrm{T}}$ denoting the steering of the qth user, and $\boldsymbol{s}(t) = [s_1(t), s_2(t), \cdots, s_Q(t)]^{\mathrm{T}} \in \mathbb{C}^Q$ denotes the signal waveform vector with ω_c representing the angular frequency of the carrier. In addition, $\boldsymbol{n}^{\mathrm{RF}}(t) \sim \mathcal{CN}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I})$ represents the zero mean additive white Gaussian noise (AWGN) vector with noise power σ_n^2.

For full digital beamforming (DBF), each antenna is connected to its own RF chain, transforming the received RF signal into a digital baseband signal. However, in a massive MIMO system with a large number of antennas, allocating separate RF chains is impractical due to hardware costs, power consumption, and computational complexity. Hybrid beamforming offers a solution that achieves comparable performance with a reduced number of RF chains.

Fig. 1 illustrates the block diagram of a fully connected hybrid beamforming structure. In this structure, a reduced number of M RF chains, where $M \ll N$, is used. The RF signal at the mth RF chain, $y_m^{\mathrm{RF}}(t)$, is obtained by analog multiplication of the input RF signal $\boldsymbol{x}^{\mathrm{RF}}(t)$ and the analog beamforming vector $\boldsymbol{\phi}_m \in \mathbb{C}^{N\times 1}$ as

$$y_m^{\mathrm{RF}}(t) = \sum_{n=1}^{N} \phi_{m,n} x_n^{\mathrm{RF}}(t), \tag{2}$$

where $\phi_{m,n}$ is the nth element of $\boldsymbol{\phi}_m$, denoting the analog beamforming coefficient for the nth antenna and the mth RF chain. The analog multiplier can be implemented using phase shifters in practice. In this case, each analog multiplier requires N phase shifters and, therefore, a total of MN phase shifters are required for implementing the F-HBF architecture. Denoting $\boldsymbol{x}(t)$ as the baseband signal of $\boldsymbol{x}^{\mathrm{RF}}(t)$, the baseband digital signal from M RF chains can be expressed as

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}\boldsymbol{x}(t) = \boldsymbol{\Phi}\boldsymbol{A}(\theta)\boldsymbol{s}(t) + \boldsymbol{n}(t), \tag{3}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_M]^{\mathrm{T}} \in \mathbb{C}^{M\times N}$ is the analog beamforming matrix.

Although the number of RF chains is reduced from the original array, the number of phase shifters may remain large for a large size array, requiring a high computational complexity to optimize. To address these issues, S-HBF can be exploited, as depicted in Fig. 2. In this configuration, the entire N-antenna array is partitioned into L non-overlapping subarrays in an identical manner. Each subarray contains $\tilde{N} = N/L$ antennas and $\tilde{M} = M/L$ RF chains, where both $\tilde{N}$ and $\tilde{M}$ are assumed to be integers. For a particular subarray, the

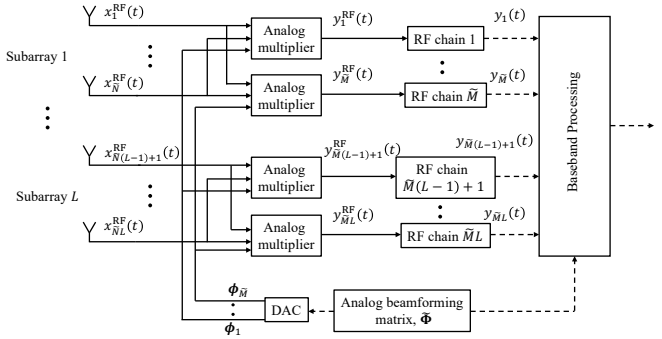


Fig. 2: Subarray-based hybrid beamforming structure.

$\tilde{M}$ RF chains are connected only to the $\tilde{N}$ antennas belonging to that subarray. Therefore, $\tilde{N}\tilde{M}L = NM/L$ phase shifters are required to implement this architecture, resulting in a reduction by a factor of L compared to the F-HBF case. Due to the reduction in the number of phase shifters, power consumption is also reduced. It is noted that, because of the identical subarray partitioning, we only need to optimize the analog beamforming matrix for a single subarray $\tilde{\boldsymbol{\Phi}}_l \in \mathbb{C}^{\tilde{M}\times\tilde{N}}$, and the result can be replicated to all subarrays. This results in a dimensional reduction by a factor of L^2 in the optimization of the analog beamforming matrix.

III. Analog Beamforming Matrix Optimization

A. Analog beamforming matrix Optimizing for uniform prior

We first describe the optimization procedure of the analog beamforming matrix $\tilde{\boldsymbol{\Phi}}_l$ for the lth subarray using an information-theoretic approach as developed in [16, 22]. In this approach, the DOA θ is considered as a random variable with a uniform prior distribution.

The probability density function (pdf) of the analog beamformer output of the lth subarray can be expressed as

$$f(\boldsymbol{y}_l) = \int_{\theta\in\Theta} f(\boldsymbol{y}_l|\theta) f(\theta) d\theta, \tag{4}$$

where $f(\theta)$ is the pdf of θ and Θ is the angular region of observation.

We discretize the observation region Θ into K angular bins with width of $\Delta\bar{\theta}$. As such, the pdf $f(\theta)$ is approximated by probability mass function (pmf), given as $p_k \approx f(\bar{\theta}_k)\Delta\bar{\theta}$ for the kth angular bin, $k = 1, 2, \cdots, K$. In this case, $f(\boldsymbol{y}_l)$ can be approximated as $f(\boldsymbol{y}_l) \approx \sum_{k=1}^{K} p_k f(\boldsymbol{y}_l|\bar{\theta}_k)$, where $f(\boldsymbol{y}_l|\bar{\theta}_k)$ is the conditional pdf of $\boldsymbol{y}_l$ given DOA $\bar{\theta}_k$. The analog beamformer output $\boldsymbol{y}_l$ for a signal impinging in the lth subarray from kth angular bin with nominal DOA $\bar{\theta}_k$ can be expressed as

$$\boldsymbol{y}_l|_{\theta=\bar{\theta}_k} = \tilde{\boldsymbol{\Phi}}_l \left[\tilde{\boldsymbol{a}}_l(\bar{\theta}_k) s(t) + \boldsymbol{n}(t)\right] \tag{5}$$

with $\tilde{\boldsymbol{a}}_l(\bar{\theta}_k)$ being the steering vector corresponding to the lth subarray and the conditional pdf

$$f(\boldsymbol{y}_l|\bar{\theta}_k) = \frac{1}{\pi^{\tilde{M}}|\tilde{\boldsymbol{C}}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k}|} e^{-\boldsymbol{y}_l^{\mathrm{H}} \tilde{\boldsymbol{C}}^{-1}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k} \boldsymbol{y}_l}, \tag{6}$$

where

$$\tilde{\boldsymbol{C}}_{\boldsymbol{y}\boldsymbol{y}|\bar{\theta}_k} = \tilde{\boldsymbol{\Phi}}_l \left(\sigma_s^2 \tilde{\boldsymbol{a}}_l(\bar{\theta}_k)\tilde{\boldsymbol{a}}_l^{\mathrm{H}}(\bar{\theta}_k) + \sigma_n^2 \boldsymbol{I}\right) \tilde{\boldsymbol{\Phi}}_l^{\mathrm{H}} = \tilde{\boldsymbol{\Phi}}_l \tilde{\boldsymbol{E}}_{l,k} \tilde{\boldsymbol{\Phi}}_l^{\mathrm{H}} \tag{7}$$

is the covariance matrix of the analog beamformer output for the lth subarray given a specific DOA $\bar{\theta}_k$, and

$$\tilde{\boldsymbol{E}}_{l,k} = \sigma_s^2 \tilde{\boldsymbol{a}}_l(\bar{\theta}_k)\tilde{\boldsymbol{a}}_l^{\mathrm{H}}(\bar{\theta}_k) + \sigma_n^2 \boldsymbol{I}. \tag{8}$$

The analog beamforming matrix $\tilde{\boldsymbol{\Phi}}_l$ is optimized by maximizing the MI between the analog beamformer output $\boldsymbol{y}_l$ and the signal DOA θ, i.e., $I(\boldsymbol{y}_l;\theta)$, in a gradient ascent manner, expressed as

$$\tilde{\boldsymbol{\Phi}}_l \leftarrow \tilde{\boldsymbol{\Phi}}_l + \gamma \nabla_{\tilde{\boldsymbol{\Phi}}_l} I(\boldsymbol{y}_l;\theta), \tag{9}$$

where $\gamma > 0$ is the step size and $\nabla_{\tilde{\boldsymbol{\Phi}}_l}$ is the gradient of the MI with respect to $\tilde{\boldsymbol{\Phi}}_l$. The MI can be expressed as

$$\begin{aligned} \boldsymbol{I}(\boldsymbol{y}_l;\theta) = &- \int f(\boldsymbol{y}_l) \log f(\boldsymbol{y}_l) d\boldsymbol{y}_l \\ &+ \int\int f(\boldsymbol{y}_l,\theta) \log f(\boldsymbol{y}_l,\theta) d\boldsymbol{y}_l d\theta. \end{aligned} \tag{10}$$

Using the approximated pmf, the gradient of the MI with respect to the analog beamforming matrix is obtained as [16]

$$\nabla_{\tilde{\boldsymbol{\Phi}}_l} I(\boldsymbol{y}_l;\theta) \approx \frac{\sum_{k=1}^{K} p_k \frac{\boldsymbol{D}_k^{-1}}{|\boldsymbol{D}_k|} \tilde{\boldsymbol{\Phi}}_l \tilde{\boldsymbol{E}}_{l,k}}{\sum_{k=1}^{K} p_k |\boldsymbol{D}_k|^{-1}} - \sum_{k=1}^{K} p_k \boldsymbol{D}_k^{-1} \tilde{\boldsymbol{\Phi}}_l \tilde{\boldsymbol{E}}_{l,k}, \tag{11}$$

where $\boldsymbol{D}_k = \sigma_n^{-2} \tilde{\boldsymbol{\Phi}}_l \tilde{\boldsymbol{E}}_{l,k} \tilde{\boldsymbol{\Phi}}_l^{\mathrm{H}}$. Note that approximate equality is used to indicate that the pmf p_k is used in lieu of the true pdf.

We then introduced a row orthonormal constraint, i.e., $\tilde{\boldsymbol{\Phi}}_l \tilde{\boldsymbol{\Phi}}_l^{\mathrm{H}} = \boldsymbol{I}$, to ensure that scaling up $\tilde{\boldsymbol{\Phi}}_l$ does not increase the MI. Each element of $\tilde{\boldsymbol{\Phi}}_l$ is normalized to have a constant magnitude for convenient implementation using phase shifters. Due to the identical subarray partitioning, we can use the same $\tilde{\boldsymbol{\Phi}}_l$ optimzied for a subarray to all other subarrays. As a result, the full analog beamforming matrix $\boldsymbol{\Phi}$ for the S-HBF settings becomes

$$\boldsymbol{\Phi} = \boldsymbol{I}_L \otimes \tilde{\boldsymbol{\Phi}}_1. \tag{12}$$

Substituting it to Eq. (3) results in the analog beamformer output of the entire array.

B. Iterative update of the DOA prior distribution

Since the pmf of DOAs p is unknown, initially it is considered as uniform distribution to optimize $\tilde{\boldsymbol{\Phi}}_l$ based on Eqs. (9) and (11), and then obtain $\boldsymbol{\Phi}$ using Eq. (12). The spatial spectrum of the user signals is then estimated by using the minimum variance distortionless response (MVDR) algorithm as

$$P^{(i)}(\theta) = \frac{1}{\tilde{N}} \frac{\boldsymbol{a}^{\mathrm{H}}(\boldsymbol{\Phi}^{(i)})^{\mathrm{H}} \boldsymbol{\Phi}^{(i)} \boldsymbol{a}(\theta)}{\boldsymbol{a}^{\mathrm{H}}(\boldsymbol{\Phi}^{(i)})^{\mathrm{H}} (\hat{\boldsymbol{R}}_{\boldsymbol{yy}}^{(i)})^{-1} \boldsymbol{\Phi}^{(i)} \boldsymbol{a}(\theta)}, \tag{13}$$

where $P^{(i)}(\theta)$ denotes the estimated spatial spectrum and $\hat{\boldsymbol{R}}_{\boldsymbol{yy}}^{(i)}$ is the sample covariance matrix of the analog beamformer output $\boldsymbol{y}^{(i)}(t)$, with the superscript $^{(i)}$ indicating the ith iteration. The normalized spatial spectrum is considered as the updated prior of the DOAs, expressed as

$$p_k^{(i+1)}(\theta) = \frac{P^{(i)}(\theta_k)}{\sum_{k=1}^{K} P^{(i)}(\theta_k)}, \tag{14}$$

which is used in the subsequent iteration for the optimization of $\boldsymbol{\Phi}$. The procedures for optimizing $\boldsymbol{\Phi}$ are summarized in Algorithm 1.

Algorithm 1: Algorithm for optimizing $\boldsymbol{\Phi}$

Input : N, M, J_{SO}, J_{SI}, $\Delta\bar{\theta}$, γ
Output: $\boldsymbol{\Phi}$

Initialize $\tilde{\boldsymbol{\Phi}}_1$ randomly.
Initialize the pmf of θ, p_k, as a uniform distribution
/* Outer loop */
for $j_{so} \leftarrow 1$ **to** J_{SO} **do**
  /* Inner loop */
  for $j_{si} \leftarrow 1$ **to** J_{SI} **do**
    Calculate the gradient of $I(\boldsymbol{y}_l;\theta)$ using Eq. (11)
    Update $\tilde{\boldsymbol{\Phi}}_1$ using (9) and make it constant modulus
  end for
  Obtain $\boldsymbol{\Phi}$ using (12)
  Perform analog beamforming as $\boldsymbol{y}(t) = \boldsymbol{\Phi}\boldsymbol{x}(t)$
  Estimating MVDR spectrum using (13)
  Update the pmf p_k as normalized MVDR spectrum using (14)
end for
return $\boldsymbol{\Phi}$

IV. Digital beamforming

After performing analog beamforming, digital beamforming is applied on the analog beamformer output to enhance the output signal-to-interference-plus-noise ratio (SINR). The MVDR beamforming weight vector for qth user is given as

$$\boldsymbol{w}_q = \frac{\hat{\boldsymbol{R}}_{i+n,q}^{-1} \boldsymbol{\Phi} \boldsymbol{a}(\hat{\theta}_q)}{\boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_q) \boldsymbol{\Phi}^{\mathrm{H}} \hat{\boldsymbol{R}}_{i+n,q}^{-1} \boldsymbol{\Phi} \boldsymbol{a}(\hat{\theta}_q)}, \tag{15}$$

where $\hat{\theta}_q$ is the estimated DOA of the qth signal,

$$\hat{\boldsymbol{R}}_{i+n,q} = \sum_{i=1, i\neq q}^{Q} P(\hat{\theta}_i) \boldsymbol{\Phi} \boldsymbol{a}(\hat{\theta}_i) \boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_i) \boldsymbol{\Phi}^{\mathrm{H}} + \hat{\sigma}_n^2 \boldsymbol{I} \tag{16}$$

is the interference-plus-noise covariance matrix for the analog beamformer output signal [16, 26], where $P(\hat{\theta}_i)$ is the estimated power, $\boldsymbol{a}(\hat{\theta}_i)$ is the estimated steering vector of the ith signal, and $\hat{\sigma}_n^2$ is the estimated noise power, which is considered as the minimum eigenvalue of $\hat{\boldsymbol{R}}_{\boldsymbol{yy}}$. Therefore, the estimation of the interference-plus-noise covariance matrix relies on the estimation of the signal DOAs and the corresponding signal power.

The signal DOAs can be estimated by finding the location of peaks of the spatial spectrum obtained using Eq. (13). Once the DOAs are estimated, the estimated steering vectors can also be subsequently obtained. On the other hand, to accurately obtain the signal power, a covariance matrix fitting problem can be formulated using the estimated DOAs as the support information as

$$\min_{\boldsymbol{P}(\hat{\boldsymbol{\theta}})} \left\| \hat{\boldsymbol{R}}_{\boldsymbol{yy}} - \boldsymbol{\Phi} \boldsymbol{A}(\hat{\boldsymbol{\theta}}) \boldsymbol{P}(\hat{\boldsymbol{\theta}}) \boldsymbol{A}^{\mathrm{H}}(\hat{\boldsymbol{\theta}}) \boldsymbol{\Phi}^{\mathrm{H}} - \sigma_n^2 \boldsymbol{I} \right\|_F^2, \tag{17}$$

where the optimization variable $\boldsymbol{P}(\hat{\boldsymbol{\theta}}) = \mathrm{Diag}([P(\hat{\theta}_1), \cdots, \hat{P}(\theta_Q)])$. The closed-form solution to Eq. (17) is given as [16, 26]

$$\boldsymbol{P}(\hat{\boldsymbol{\theta}}) = \left[\boldsymbol{G}^{\mathrm{H}} \boldsymbol{G} \right]^{-1} \boldsymbol{G}^{\mathrm{H}} \boldsymbol{z}, \tag{18}$$

where $\boldsymbol{G}$ is an $M^2 \times Q$ matrix given as

$$\boldsymbol{G} = \left[\mathrm{vec}(\boldsymbol{\Phi} \boldsymbol{a}(\hat{\theta}_1) \boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_1) \boldsymbol{\Phi}^{\mathrm{H}}), \cdots, \mathrm{vec}(\boldsymbol{\Phi} \boldsymbol{a}(\hat{\theta}_Q) \boldsymbol{a}^{\mathrm{H}}(\hat{\theta}_Q) \boldsymbol{\Phi}^{\mathrm{H}}) \right], \tag{19}$$

and $\boldsymbol{z} = \mathrm{vec}(\hat{\boldsymbol{R}}_{\boldsymbol{yy}} - \hat{\sigma}_n^2 \boldsymbol{I})$ is a vector of dimension M^2. Once the estimated DOAs and the corresponding powers are obtained, the digital beamforming weights can be obtained from Eq. (15).

V. Analysis of Computational Complexity

The S-HBF can lower the hardware requirements for implementing a massive MIMO system and the computational complexity for optimizing the analog beamforming matrix.

The complexity of computing the gradient of MI for the F-HBF requires a complexity of $\mathcal{O}(KMN^2)$ [16]. Considering J_{FI} inner iterations are required to optimize $\boldsymbol{\Phi}$ for a particular DOA prior

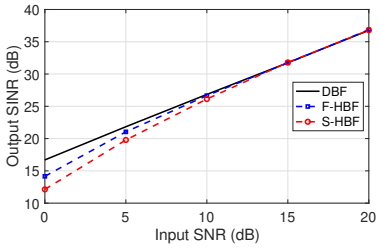


Fig. 3: Output SINR versus input SNR.

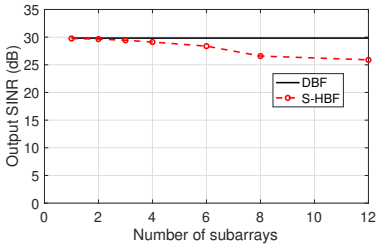


Fig. 4: Output SINR versus number of subarrays.

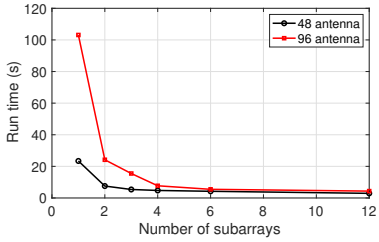
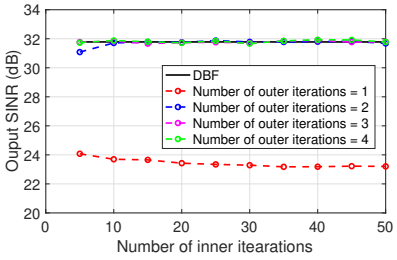
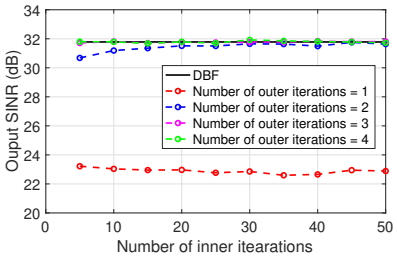


Fig. 5: Runtime for optimizing $\mathbf{\Phi}$ versus the number of subarrays.



(a) F-HBF



(b) S-HBF

Fig. 6: Output SINR versus number of iterations.

and J_{FO} outer iterations are used to reoptimize $\mathbf{\Phi}$ by updating the prior of DOAs, the total computational complexity for optimizing $\mathbf{\Phi}$ is $\mathcal{O}(J_{FO}J_{FI}KMN^2)$. On the other hand, for the S-HBF structure, the computational complexity required for optimizing $\mathbf{\Phi}$ is $\mathcal{O}(J_{SO}J_{SI}K\tilde{M}\tilde{N}^2) = O(J_{SO}J_{SI}KMN^2/L^3)$, which reduces the complexity by a factor of L^3. Note that the complexity is proportional to J_{SO} and J_{SI}, which respectively denote the numbers of outer and inner iterations required for the S-HBF case.

VI. Simulation Results

We compare the performance of the DBF, F-HBF, and S-HBF in terms of the output SINR performance. We consider a uniform linear receive array comprising $N = 48$ antennas with a half-wavelength inter-element spacing in a massive MIMO system, and $M = 12$ RF chains are used in the hybrid beamforming structures. The pdf of the DOAs are discretized with an width of $\Delta\bar{\theta} = 0.01\,\text{deg}$, resulting in a total $K = 1801$ grids. The DOAs of the desired source and two interfering users are chosen from a uniform distribution with the range between -60° to 60°. In the iterative optimization of the analog beamforming matrix $\mathbf{\Phi}$ based on gradient ascent update, the step size is chosen as $\gamma = 0.001$. We perform 20 inner iterations and 3 outer iterations to optimize $\mathbf{\Phi}$.

In the first example, we consider the desired user with DOA 5° with a varying input signal-to-noise ratio (SNR), whereas two interfering users have DOAs of -6° and 15° and an input interference-to-noise ratio (INR) of 10 dB, and 100 snapshots are considered. For the S-HBF case, $L = 2$ subarrays are assumed. Fig. 3 depicts the output SINR for DBF, F-HBF, and S-HBF with respect to the input SNR. It demonstrates that DBF and F-HBF result in equivalent performance. For S-HBF, there is a noticeable performance loss at a low input SNR, but such loss vanishes as the input SNR increases. For example, the output SINR differs by approximately 4 dB at 0 dB input SNR, whereas no difference is observed when the input SNR is 10 dB. Note that, for the F-HBF case, $MN = 500$ phase shifters are required, while for S-HBF, the number of required phase shifters is reduced to $MN/L = 250$.

Fig. 4 shows the output SINR with respect to the number of subarrays in an S-HBF structure. For convenience of subarray division, we consider $N = 96$ antennas, $M = 24$ RF chains, and different values of $L = 1, 2, 3, 4, 6, 8, 12$ subarrays. The input SNR is set at 10 dB. Note that, when the number of subarrays is $L = 1$, the S-HBF becomes equivalent to the F-HBF. It can be observed that, when the number of subarrays is $L = 2$ or $L = 3$, the output SINR performance of the S-HBF is very close to that of the F-HBF, whereas approximately 3 dB of SINR loss is observed when $L = 4$.

As discussed previously, the S-HBF structure reduces computational complexity for the analog beamformer optimization. Fig. 5 depicts the MATLAB runtime with respect to the number of subarrays L for two S-HBF architectures, one with 48 antennas and 12 RF chains and the other with 96 antennas and 24 RF chains. It is observed that, even with only $L = 2$ subarrays, the S-HBF structure reduces the runtime significantly compared to the F-HBF case.

Fig. 6 compares the output SINR performance of F-HBF and S-HBF across various numbers of inner and outer iterations for an input SNR of 15 dB. For a single outer iteration, both F-HBF and S-HBF converge to only about 73% of the accuracy of the fully digital case. However, as the number of outer iterations increases, the convergence accuracy improves rapidly. With 2 outer iterations, only 5 inner iterations are needed to achieve 97.86% accuracy for the F-HBF case, and around 10 inner iterations are required for full convergence. In contrast, for the S-HBF case, 5 inner iterations are needed to reach 96.54% accuracy, while 45 inner iterations are required for full convergence. For three or more outer iterations, both F-HBF and S-HBF achieve full accuracy with a small number of inner iterations (e.g., 5), demonstrating the fast convergence of the optimization method in both cases.

VII. Conclusion

In this paper, we presented an efficient solution for the implementation of a hybrid beamforming structure in a massive MIMO system. This subarray-based hybrid beamforming architecture offers a hardware- and power-efficient solution. By sharing the analog beamforming weights across the subarrays, the complexity of analog beamformer optimization is also significantly reduced. Simulation results verify the effectiveness of the proposed approach in terms of the output SINR and optimization complexity.

References

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2012.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[4] F. Jiang, J. Chen, A. L. Swindlehurst, and J. A. López-Salcedo, "Massive MIMO for wireless sensing with a coherent multiple access channel," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3005–3017, 2015.

[5] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.

[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] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, 2014.

[8] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, 2017.

[9] Y. Gu and Y. D. Zhang, "Information-theoretic pilot design for downlink channel estimation in FDD massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2334–2346, 2019.

[10] Y. Gu, Y. D. Zhang, and N. A. Goodman, "Optimized compressive sensing-based direction-of-arrival estimation in massive MIMO," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, New Orleans, LA, March 2017, pp. 3181–3185.

[11] X. Wu, D. Liu, and F. Yin, "Hybrid beamforming for multi-user massive mimo systems," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3879–3891, 2018.

[12] M. Guo, Y. D. Zhang, and T. Chen, "DOA estimation using compressed sparse array," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4133–4146, 2018.

[13] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 4, pp. 3060–3097, 2018.

[14] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. B. Letaief, "Hybrid beamforming for millimeter wave systems using the MMSE criterion," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3693–3708, 2019.

[15] J. Zhang, X. Yu, and K. B. Letaief, "Hybrid beamforming for 5G and beyond millimeter-wave systems: A holistic view," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 77–91, 2019.

[16] Y. Gu and Y. D. Zhang, "Compressive sampling optimization for user signal parameter estimation in massive MIMO systems," *Digital Signal Process.*, vol. 94, pp. 105–113, 2019.

[17] S. R. Pavel, Y. D. Zhang, M. S. Greco, and F. Gini, "Deep learning-based compressive sampling optimization in massive MIMO systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Rhodes Island, Greece, June 2023, pp. 1–5.

[18] S. R. Pavel and Y. D. Zhang, "Optimization of the compressive measurement matrix in a massive MIMO system exploiting LSTM networks," *Algorithms*, vol. 16, no. 261, pp. 1–16, 2023.

[19] J. Zhang, S. Li, L. Jin, W. Liu, and H. C. So, "Multi-beam multiplexing design with phase-only excitation based on hybrid beamforming architectures," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Seoul, Korea, April 2024, pp. 8451–8455.

[20] B. K. Chalise and M. G. Amin, "Hybrid MVDR beamformer for subarray architecture with sparse recovery and manifold optimization methods," in *Proc. IEEE Radar Conf.*, Denver, CO, May 2024, pp. 1–6.

[21] Y. D. Zhang and S. Sun, "Identical partitioning of consecutive integer set," in *Proc. IEEE Sensor Array and Multich. Signal Process. Workshop*, Corvallis, OR, 2024.

[22] Y. D. Zhang, "Iterative learning for optimized compressive measurements in massive MIMO systems," in *Proc. IEEE Radar Conf.*, New York, NY, 2022, pp. 1–5.

[23] M. S. R. Pavel and Y. D. Zhang, "Deep learning-based robust imaging exploiting 2-D array compressive measurement," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 2022, pp. 1433–1437.

[24] F. Sohrabi, Z. Chen, and W. Yu, "Deep active learning approach to adaptive beamforming for mmWave initial alignment," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2347–2360, 2021.

[25] S. R. Pavel and Y. D. Zhang, "Neural network approach to iterative optimization of compressive measurement matrix in massive MIMO system," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Trondheim, Norway, June 2022, pp. 171–175.

[26] Y. Gu, N. A. Goodman, and Y. D. Zhang, "Adaptive beamforming via sparsity-based reconstruction of covariance matrix," in *Compressed Sensing in Radar Signal Processing*, A. De Maio, Y. Eldar, and A. Haimovich, Eds. Cambridge Univ. Press, 2019.