Beamspace MIMO Systems With Reduced Beam Selection Complexity

Sinasi Cetinkaya, *Student Member, IEEE* and Hüseyin Arslan, *Fellow, IEEE*

***Abstract*—Beamspace multiple-input multiple-output (B-MIMO) systems with proper beam selection (BS) promise to lower the radio-frequency (RF) chains overhead with no noticeable degradation in performance. Most of the existing BS schemes are not practical due to the computational cost arising from iterative search or alternating optimization. Hence, this letter examines the complexity reduction of the existing BS with incremental QR precoder (I-QR-P) and decremental QR precoder (D-QR-P). The proposed two-stage and three-stage algorithms reduce the complexity of D-QR-P and I-QR-P, respectively. Both aim to lower complexity by decreasing the candidate beam size by eliminating the beams with no contribution to any user and using matrix perturbation theory to update QR decompositions. Numerical results reveal that the proposed algorithms considerably reduce the complexity while maintaining a similar sum-rate with baseline algorithms.**

***Index Terms*—Beamspace MIMO, beam selection, mmWave communication, precoding, QR decomposition.**

I. Introduction

UTILIZATION of the millimeter-wave (mmWave) spectrum is presumed to be a key enabler for emerging next-generation wireless communication networks [1]. Since mmWave frequencies offer small wavelengths, it is possible to pack many antennas into small physical areas. This property enables a promising marriage between mmWave frequencies and massive multiple-input multiple-output (M-MIMO), thereby overcoming the severe free-space path-loss due to high directional beamforming gain [2]. MmWave M-MIMO also enhances spectral efficiency by allowing multiple data streams [3] and utilizing its larger bandwidth [4]. However, excessive power consumption and hardware cost are its drawbacks since each antenna entails its own RF chain. Beamspace MIMO (B-MIMO) proposed in [5] has, therefore, received much consideration to reduce the RF chain requirement by taking advantage of the inherent sparsity in the mmWave channels. In B-MIMO, the angular domain (i.e., beamspace) representation of the spatial channel is performed by employing a discrete lens array (DLA) at the base station to explore the channel sparsity [5]. Hence, a reduction in radio-frequency (RF) chains required is achieved without compromising the system performance by performing beam selection (BS) [4], [5].

Magnitude maximization BS (MM-BS) [5] assigns a beam proving the maximum received signal power to each user. Nevertheless, it suffers from high multi-user interference and RF redundancy; however, these limitations were handled by interference-aware BS (IA-BS) [4]. Several signal-to-interference-plus-noise ratio (SINR) and sum-rate maximization-based iterative BS algorithms were investigated in [6]–[8], while zero-forcing (ZF) precoding was employed to eliminate the multi-user interference. By enabling multiple-beam group selection in [9], rate-loss was mitigated by creating a reliable channel cluster for each user. Recently, BS and precoding schemes were studied for a new B-MIMO architecture with lens antenna subarrays (LASs) [10]. Several studies [11], [12] on BS for wideband B-MIMO were also conducted to overcome the beam squint occurring in mmWave.

Due to the limited power at the base stations and to avoid latency, the complexity of several BS methods has been investigated. In [13], low-complexity BS methods were proposed based on the graph theory and heuristic greedy algorithm. QR decomposition of the beamspace channel was inspected in [14], and an iterative BS algorithm and a precoder for eliminating multi-user interference were proposed. Along with its outstanding system performance over the existing algorithms, it is not practical due to its high complexity. In [15], the complexity of the conventional ZF precoder and QR precoder [14] were probed and reduced by updating the factorization or decomposition results using matrix perturbation theory instead of performing from scratch again.

This letter proposes two complexity-reduced BS algorithms. The main contributions are summarized as follows:

- The proposed two-stage and three-stage BS algorithms remarkably decrease the complexity of D-QR precoder (D-QR-P) [15] and I-QR precoder (I-QR-P) [15], respectively.
- Both enjoy considerably higher sum-rate than the existing algorithms proposed in [4]–[6].
- The sum-rate performance of D-QR-P is high; however, it suffers from high complexity as the number of antennas increases. The two-stage BS obtains the complexity reduction with almost identical sum-rate performance.
- The three-stage BS algorithm significantly reduces complexity while compensating for the sum-rate performance loss at low signal-to-noise ratio (SNR)s that I-QR-P suffers from.
- Both utilize the matrix perturbation to update QR decomposition and aim to decrease the beam size by removing the beams with no contribution to any user from the beam set.

Notation: $\mathbf{A}$, $\mathbf{a}$, a, $\mathcal{A}$ denote a matrix, a vector, a scalar, and a set, respectively. $\mathbf{A}^H$, $\mathbf{A}^T$, $\mathbf{A}^{-1}$ are Hermitian, transpose, and inverse of $\mathbf{A}$ respectively. $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix with $\mathbf{a}$ on its diagonal. $\mathbf{I}$ is the identity matrix, and $\mathbb{C}^{M\times N}$ is the space of $M \times N$ complex-valued matrices, $\mathbb{E}[\cdot]$ is the expectation operator, and $\mathrm{Card}(\mathcal{A})$ is the cardinality of $\mathcal{A}$.

S. Cetinkaya is with the Department of Electrical Engineering, University of South Florida, Tampa, FL, 33620, USA (e-mail: scetinkaya@usf.edu).

H. Arslan is with the Department of Electrical and Electronics Engineering, Istanbul Medipol University, Istanbul, 34810, Turkey (e-mail: huseyinarslan@medipol.edu.tr).

II. System Model

This paper considers a B-MIMO architecture in a downlink mmWave scenario. For a conventional M-MIMO architecture where a base station consists of N antennas modeled as a uniform linear array (ULA) to serve K single-antenna users, the received signal vector $\mathbf{y} \in \mathbb{C}^{K\times 1}$ is expressed as

$$\mathbf{y} = \mathbf{H}^H\mathbf{P}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N\times K}$ stands for the channel matrix where $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ denotes the channel vector between k-th user and the base station. The normalized transmitted signal vector is defined by $\mathbf{s} \in \mathbb{C}^{K\times 1}$ fulfilling $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_K$ and $\mathbf{P} \in \mathbb{C}^{N\times K}$ is the precoding matrix designed to cancel the multi-user interference. Additionally, $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_K)$ is additive white Gaussian noise (AWGN).

This paper considers the Saleh-Valenzuela mmWave channel model [16], which models $\mathbf{h}_k$ as [4]

$$\mathbf{h}_k = \sum_{l=0}^{L} \alpha_k^{(l)}\mathbf{a}\left(\varphi_k^{(l)}\right). \tag{2}$$

In (2), the complex gain and the spatial direction of the k-th user for the l-th path are stated by $\alpha_k^{(l)}$ and $\varphi_k^{(l)}$, respectively. Note that $l = 0$ refers to the line-of-sight (LoS) component, while $l = 1, 2, ..., L$ represents the non-line-of-sight (NLoS) components. For the N-element typical ULA, the array steering vector of the l-th path is stated as [4] $\mathbf{a}(\varphi) = \frac{1}{\sqrt{N}}\left[e^{-j2\pi\varphi b}\right]_{b\in\mathcal{I}(N)} \in \mathbb{C}^{N\times 1}$, where $\mathcal{I}(N) = \{p - (N-1)/2, p = 0, 1, ..., N-1\}$ is a symetric set of indices centered around zero. Moreover, $\varphi = \frac{d}{\lambda}\sin(\theta)$, where θ, $d = \lambda/2$, and λ denote the physical direction, the antenna spacing, and the carrier signal wavelength, respectively.

The mmWave channel is inherently sparse since the LoS components of the channel strongly dominate the NLoS components [5]. The use of a DLA at the base station converts the spatial channel (2) into the beamspace channel as it behaves like a spatial discrete Fourier transformer represented by matrix $\mathbf{U} \in \mathbb{C}^{N\times N}$. Specifically, $\mathbf{U} = [\mathbf{a}(\bar{\varphi}_1), \mathbf{a}(\bar{\varphi}_2), \cdots, \mathbf{a}(\bar{\varphi}_N)]^H$ consists of the array steering vectors corresponding to N predefined orthogonal directions covering the entire angular space [4], [15], where $\bar{\varphi_n} = \frac{1}{N}(n - \frac{N+1}{2})$ for $n = 1, 2, \ldots, N$ stands for the predefined spatial directions. Ultimately, the beamspace channel[1] is $\mathbf{H}_b = \mathbf{U}\mathbf{H}$, and the corresponding received signal vector $\mathbf{y}_b$ is given as

$$\mathbf{y}_b = {\mathbf{H}_b}^H\mathbf{P}\mathbf{s} + \mathbf{n}. \tag{3}$$

The beamspace channel can be represented with a considerably small number of precisely chosen beams without compromising the system performance due to intrinsic sparsity. As a result of the BS process, the dimension-reduced M-MIMO, so-called B-MIMO, is obtained as [4]

$$\mathbf{y}_b \approx \tilde{\mathbf{H}}_r^H\mathbf{P}_r\mathbf{s} + \mathbf{n}, \tag{4}$$

where the dimension-reduced beamspace channel is $\tilde{\mathbf{H}}_\mathbf{r} = \mathbf{H}_\mathbf{b}(\mathcal{B}, :)$, in which $\mathcal{B}$ denotes the set containing the chosen beam indexes, and $\mathbf{P}_r$ stands for the corresponding dimension-reduced precoding matrix.

[1] We assume that channel state information (CSI) is known by the base station where channel estimation can be performed as in [17].

Note that using DLA and performing BS reduces the required RF chains while preserving the narrow beamwidth [6]. As a result, the adopted B-MIMO architecture is suitable for mmWave systems due to low hardware complexity and high antenna gain properties, even with fewer RF chains [6].

Additionaly, the base station communicates with every user in set $\mathcal{K} = \{1, 2, \ldots, K\}$ via *only one data stream* to assure the spatial multiplexing gain. Thus, the number of data streams and RF chains are set to $N_s = K$ and $N_{RF} = K$, respectively.

A. QR Decomposition of Dimension-Reduced B-MIMO

Let $\tilde{\mathbf{H}}_r$ be decomposed into a unitary matrix of $\mathbf{Q} \in \mathbb{C}^{K\times K}$ and an upper triangular matrix of $\mathbf{R} \in \mathbb{C}^{K\times K}$, such that $\tilde{\mathbf{H}}_r = \mathbf{Q}\mathbf{R}$ [18]. Hence, (4) becomes $\mathbf{y}_b \approx \mathbf{R}^H\mathbf{s} + \mathbf{n}$ when the precoder is $\mathbf{P}_r = \mathbf{Q}$. Thus, the k-th user receives [14]

$$\tilde{y}_k = \tilde{r}_{kk}s_k + I_k + n_k, \tag{5}$$

where $\tilde{r}_{kk}$ equals to the k-th element of $\mathrm{diag}(\mathbf{R})$, and the interference $I_k = \sum_{k>j}\tilde{r}_{kj}s_j$ can be eliminated for all users by diagonalizing $\mathbf{R}^H$. Let the precoder be $\mathbf{P}_r = \mathbf{Q}\mathbf{G}$, where $\mathbf{G} \in \mathbb{C}^{K\times K}$ is the Given rotations such that the diagonal elements of $\mathbf{R}^H\mathbf{G} \in \mathbb{C}^{K\times K}$ and $\mathbf{R}^H$ are same. Accordingly, the sum-rate is [14], [15]

$$R_{sum} = \sum_k \log_2\left(1 + \frac{\gamma}{K}\tilde{r}_{kk}^2\right) \quad \text{bit/s/Hz}, \tag{6}$$

where ρ and $\gamma = \rho/\sigma^2$ stand for the signal power and SNR, respectively.

B. QR Decomposition Update

Matrix perturbation theory allows QR decomposition of a matrix to be updated easily instead of recomputing from scratch when the matrix undergone a modification. Suppose we have the decomposition of $\mathbf{H}_b = \mathbf{Q}\mathbf{R}$, and let $\mathbf{H}_b^{(\pm n)} = \mathbf{H}_b - \mathbf{u}\mathbf{z}^T$ represent the modified matrix after inserting or eliminating the n-th row (i.e., $\mathbf{z}^T$), where $\mathbf{u} = \mp\mathbf{e}_n$. Note that QR decomposition is called incremental QR (I-QR) decomposition when a new row inserted to a matrix, while it is called decremental QR (D-QR) decomposition when a row is deleted from a matrix. The upper-Hessenberg matrix of $\mathbf{H}_b^{(\pm n)}$ can be expressed as $\mathbf{H}_b^{(\pm n)} = \mathbf{Q}[\mathbf{R} + \mathbf{w}\mathbf{z}^T]$, where $\mathbf{w} = \mathbf{Q}^H\mathbf{u}$. Denoting $\mathbf{J}_m$ is a Given rotation acting in planes m and $m+1$, where $m = 1, 2, \ldots, (N-1)$, the series of rotations is obtained by $\mathbf{J}_1^T \ldots \mathbf{J}_{N-1}^T\mathbf{w} = \mp||w||_2\mathbf{e}_1$, where $\mathbf{e}_1 = (1, 0, \ldots, 0)$ represents the unit vector. Assuming that same rotations are applied to $\mathbf{R}$, we acquire an upper-Hessenber matrix of $\mathbf{H}_0 = \mathbf{J}_1^T \ldots \mathbf{J}_{N-1}^T\mathbf{R}$. Ultimately, $\mathbf{H}_1 = \mathbf{J}_1^T \ldots \mathbf{J}_{N-1}^T[\mathbf{R} + \mathbf{w}\mathbf{z}^T] = \mathbf{H}_0 \mp ||w||_2\mathbf{e}_1\mathbf{z}^T$. To update $\mathbf{R}^{(\pm n)}$, $(N-1)$ $\mathbf{G}_m$ Given rotations are applied to $\mathbf{H}_1$ [18] such that

$$\mathbf{R}^{(\pm n)} = \mathbf{G}_{N-1}^T\mathbf{G}_{N-2}^T \ldots \mathbf{G}_1^T\mathbf{H}_1, \tag{7}$$

is an upper triangular matrix. Then, $\mathbf{Q}$ can be updated as [18]

$$\mathbf{Q}^{(\pm n)} = \mathbf{Q}\mathbf{J}_{N-1} \ldots \mathbf{J}_1\mathbf{G}_1 \ldots \mathbf{G}_{N-1}. \tag{8}$$

Note that, we omit the values of rotation matrices $\mathbf{J}_m$ and $\mathbf{G}_m$; however, details are available in [18].

III. Proposed Beam Selection Algorithms

This section revisits the I-QR-P and D-QR-P given in [15], and investigates the complexity of BS for further reduction.

A. Proposed Two Stage Beam Selection with D-QR-P

Conventional QR precoder (C-QR-P) [14] for BS is an iterative process where beams with a minimum contribution to system performance (i.e., sum-rate) are discarded, causing unaffordable computational complexity due to the required QR decomposition from scratch in each iteration such that the number of required iterations in the outer loop is $(N-K)$ in the i-th iteration, where $i=0,1,\ldots,(N-K-1)$. However, the main complexity arises from the inner loop which contains $(N-i)$ QR decomposition operation to eliminate the beam with the least contribution to the system sum-rate from the beam set. Since the complexity to compute the QR decomposition from scratch is $\mathcal{O}(2(N-i)K^2)$, the total complexity of C-QR-P is $\sum_{i=0}^{N-K-1}(N-i)\mathcal{O}(2(N-i)K^2) = \mathcal{O}((2K^2N^3-2K^5)/3)$ [14].

To overcome the complexity problem of C-QR-P, D-QR-P [15] was proposed by utilizing matrix perturbation theory. Note that D-QR-P updates $\mathbf{R}$ and $\mathbf{Q}$ using (7) and (8), respectively while D-QR-P regenerates from scratch when a row is deleted. The QR update process can be executed in $\mathcal{O}(4K(N-i)+4K^2)$ in the i-th iteration. Thus, the BS with D-QR-P requires the complexity of $\sum_{i=0}^{N-K-1}(N-i)\mathcal{O}(4K(N-i)+4K^2) = \mathcal{O}((4KN^3+6K^2N^2-10K^4)/3)$ [15]. Although it significantly reduces the complexity while providing almost similar sum-rate performance with [14], it is still not practical, especially when $N \geq K$ is large [15]. Therefore, we propose Algorithm 1 which consists of the following two stages.

Algorithm 1: Two Stage Beam Selection with D-QR-P

Input: $\mathbf{H}_b$, $\mathcal{D}$, $\mathcal{K}$, $\mathcal{G}=\emptyset$, $\mathcal{B}=\emptyset$, $M=0$
Output: $\tilde{\mathbf{H}}_r$

Stage 1: Identify M strongest beams for all users
while Card$(\mathcal{G}) < K$ **do**
  $M = M+1$,
  for $l = 1:K$ **do**
    $\mathbf{t}$ = sort$(|\mathbf{H}_b(:,l)|)$, and $\mathcal{G} = \mathcal{G} \cup \{\mathcal{D}(\mathbf{t}(1:M))\}$,
  end
end
Stage 2: Beam Selection with D-QR-P
$\mathbf{A} = \mathbf{H}_b(\mathcal{G},:)$, and $\mathbf{A} = \mathbf{QR}$,
for $j = 0:$ Card$(\mathcal{G}) - K - 1$ **do**
  for $k = 1:$ Card$(\mathcal{G}) - j$ **do**
    $\mathbf{u} = \mathbf{e}_k$, and $\mathbf{z} = \mathbf{A}(k,:)$,
    Update $\mathbf{R}^{(-k)}$, $\mathbf{Q}^{(-k)}$, and obtain $R_{sum}^{(k)}$ using (6),
  end
  $b_j = \underset{k}{\text{argmax}}\{R_{sum}^{(k)}\}$, $\mathcal{G} = \mathcal{G}\backslash\{b_i\}$, $\mathcal{B} = \mathcal{B} \cup b_j$,
  $\mathbf{A} = \mathbf{A}(\mathcal{G},:)$, $\mathbf{R} = \mathbf{R}^{(-b_j)}$, and $\mathbf{Q} = \mathbf{Q}^{(-b_j)}$,
end
$\tilde{\mathbf{H}}_r = \mathbf{H}_b(\mathcal{B},:)$

1) Identify M Strongest Beams for All K Users: This stage aims to reduce the number of QR updates in the inner loop of the D-QR-P by decreasing the number of candidate beams. Let's consider the following two definitions for lucidity.

Definition 1: Let $b_{k,m}$ represent the m-th strongest beam of the k-th user, the strongest beam $b_{k,1} \in \mathcal{D}$ contains the most of the channel power, and it is the first element of the sorted $|\mathbf{H}_b(:,k)|$ in descending order. Then, $\mathcal{G}_k^* = \{b_{k,m}\}_{m=1}^{M} \in \mathcal{D}$ includes the indices for the M strongest beams, where $\mathcal{D} = \{1,2,\ldots,N\}$ is the set containing all beams available.

Definition 2: Users sharing identical beams are called interfering user (IU)s, while a user is defined as non-interfering user (NIU) if its beam is not selected by any other users. The sets representing the IUs and NIUs are defined by $\mathcal{K}_{IU}$ and $\mathcal{K}_{NIU}$, respectively.

In this stage, the algorithm first identifies $\mathcal{G}_k^*$ for all K users. Then, the candidate beam set is $\mathcal{G} = \mathcal{G}_1^* \cup \mathcal{G}_2^* \cup \cdots \cup \mathcal{G}_K^*$, where $\mathcal{G} \subset \mathcal{D}$ and Card$(\mathcal{G}) \leq MK$. Note that this stage decides the value of M to provide enough beam diversity for all K users so that they can be served by K best-unshared beams simultaneously. An example of how the value of M is decided is shown in Fig. 1. In Fig. 1 (a) $\mathcal{G} = \{1,3,8\}$ when $M=1$. Since Card$(\mathcal{G}) < K$, there are not enough beams for all K users. Therefore, the algorithm tries the case of $M=2$ in Fig. 1 (b), where $\mathcal{G} = \{1,3,7,8\}$ and still Card$(\mathcal{G}) < K$. Thence, the case of $M=3$ is tried as in Fig. 1 (c), where $\mathcal{G} = \{1,2,3,4,7,8\}$ and Card$(\mathcal{G}) > K$. There is now enough beam diversity to provide an unshared beam for all K users. Since beam 5 and 6 have no contribution to non of the users, they are removed from the candidate beam set as in Fig. 1d.

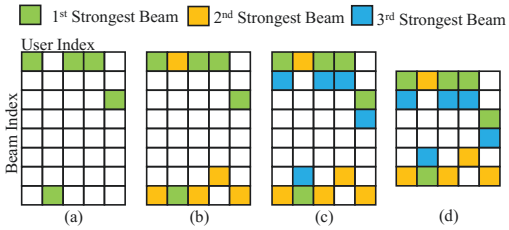

Fig. 1: An example of showing the decision of M (a) $M=1$, (b) $M=2$, (c) $M=3$, (d) Candidate beam set.

If this stage decides that $M=1$, there are no IUs in the network. In this case, Algorithm 1 selects the K strongest beams as in *Definition 1*, resulting in a near-optimal solution, already proven in [4]. Additionally, multi-user interference is eliminated by the precoder $\mathbf{P}_r = \mathbf{Q}$ obtained in step 9.

2) Beam Selection with D-QR-P: If $M > 1$, Algorithm 1 performs the BS process summarized in stage 2. Similar to the D-QR-P in [15], the beam that contributes the least to the sum-rate performance is discarded in each iteration. However, the number of $\mathbf{Q}$ and $\mathbf{R}$ updates in the inner loop is reduced from $(N-i)$ to $(\text{Card}(\mathcal{G})-i)$. Thence, the overall complexity of Algorithm 1 is $\mathcal{O}(2\text{Card}(\mathcal{G})K^2)$(step 9)$+\mathcal{O}((4K(\text{Card}(\mathcal{G}))^3 + 6K^2(\text{Card}(\mathcal{G}))^2 - 10K^4)/3)$(complexity of D-QR updates).

B. Proposed Three Stage Beam Selection with I-QR-P

We propose Algorithm 2 to reduce the complexity of I-QR-P [15]. It consists of the following three stages.

1) Beam Selection for NIUs: The complexity of the I-QR-P arises from the search process to identify the first beam that contributes the most to the system sum-rate since it requires complex QR decomposition for all N beams. To avoid this process, Algorithm 2 groups users as IUs and NIUs inspired by [4] in this stage. To do this, it first identifies the strongest beam set $\mathcal{B} = \{b_1, b_2, \ldots, b_K\}$ for all K user as in *Definition 1*. Since the probability of having IUs, $P = 1 - \frac{N!}{N^K(N-K)!}$, is considerably high in spite of N being large [4], it then removes the repeatedly selected beams from $\mathcal{B}$ and defines the non-interfering beam set as $\mathcal{V} \subset \mathcal{D}$. Since the beams in set $\mathcal{V}$ contain the most of the channel power and cause considerably low interference to others, the algorithm directly assigns these beams to the NIUs. Note that no beams are yet selected for the IUs in this stage.

2) Identify M Strongest Beams for the IUs: After NIUs are directly assigned with the beams in $\mathcal{V}$, we have a beamspace channel matrix $\mathbf{A} = \mathbf{H}_b(\mathcal{V}, :) \in \mathbb{C}^{\text{Card}(\mathcal{V}) \times K}$ for the NIUs and candidate beam set for IUs are updated as $\mathcal{D} = \mathcal{D} \backslash \mathcal{V}$, where $\text{Card}(\mathcal{D}) = N - \text{Card}(\mathcal{V})$. This stage aims to reduce the size of the candidate beams for the IUs. Let $\mathbf{B} = \mathbf{H}_b(\mathcal{D}, \mathcal{K}_{IU}) \in \mathbb{C}^{\text{Card}(\mathcal{D}) \times \text{Card}(\mathcal{K}_{\mathcal{IU}})}$ represent the beamspace channel for the IUs. Following the same process presented in Section III-A1, Algorithm 2 decides the value of M and acquires a new beam set for IUs as $\mathcal{G} \subset \mathcal{D}$, where $\text{Card}(\mathcal{G}) \leq M\text{Card}(\mathcal{K}_{IU})$.

Algorithm 2: Three Stage Beam Selection with I-QR-P

Input: $\mathbf{H}_b$, $\mathcal{D}$, $\mathcal{K}$, $\mathcal{K}_{IU} = \mathcal{K}_{NIU} = \mathcal{G} = \mathcal{B} = \emptyset$, $M = 0$
Output: $\tilde{\mathbf{H}}_r$

Stage 1: Beam Selection for NIUs
for $k = 1 : K$ **do**
  $b_k = \text{argmax}_{b_k} |\mathbf{H}_b(:, k)|$, and $\mathcal{B} = \mathcal{B} \cup \{b_k\}$,
end
Set $\mathcal{V} = \text{unique}(\mathcal{B})$,
for $i = 1 : \text{Card}(\mathcal{V})$ **do**
  $j = \text{find}(\mathcal{B} == \mathcal{V}(j))$,
  if $\text{Card}(j) > 1$ **then**
    $\mathcal{V}(\mathcal{V} == \mathcal{V}(j)) = \emptyset$, and $\mathcal{K}_{IU} = \mathcal{K}_{IU} \cup \{j\}$,
  end
end
$\mathcal{B} = \mathcal{V}$, $\mathcal{K}_{NIU} = \mathcal{K} \backslash \mathcal{K}_{IU}$, $\mathcal{D} = \mathcal{D} \backslash \mathcal{B}$, $\mathbf{A} = \mathbf{H}_b(\mathcal{B}, :)$, and $\mathbf{B} = \mathbf{H}_b(\mathcal{D}, \mathcal{K}_{IU})$,
Stage 2: Identify M strongest beams for the IUs
Follow same steps (2 to 7) in **Algorithm 1** to obtain $\mathcal{G}$
Stage 3: Beam Selection with I-QR-P for IUs
$\mathbf{A} = \mathbf{QR}$,
for $i = 1 : \text{Card}(\mathcal{K}_{IN})$ **do**
  for $j = 1 : \text{Card}(\mathcal{G})$ **do**
    $\mathbf{u} = \mathbf{e}_j$, and $\mathbf{z} = \mathbf{B}(j, :)$,
    Update $\mathbf{R}^{(+j)}$, $\mathbf{Q}^{(+j)}$, obtain $R_{sum}^{(j)}$ using (9),
  end
  $b_i = \underset{j}{\text{argmax}}\{R_{sum}^{(j)}\}$, $\mathcal{B} = \mathcal{B} \cup \{b_i\}$, $\mathcal{G} = \mathcal{G} \backslash \{b_i\}$,
  $\mathbf{Q} = \mathbf{Q}^{b_i}$, and $\mathbf{R} = \mathbf{R}^{b_i}$,
end
$\tilde{\mathbf{H}}_r = \mathbf{H}_b(\mathcal{B}, :)$

3) Beam Selection for the IUs with I-QR-P: Algorithm 2 overcomes the computational complexity of I-QR-P [15] mentioned in Section III-B1 since it does not include this step. Instead, it first decomposes the channel matrix $\mathbf{A} = \mathbf{QR}$ obtained for the NIUs in stage 1, then keeps adding a new row to $\mathbf{A}$ (i.e., a new beam) from $\mathcal{G}$ iteratively until all IUs have an unshared beam, and $\mathbf{R}$ and $\mathbf{Q}$ are updated using (7) and (8), respectively. Since the number of selected beams K' is less than K with I-QR-P BS, (6) is modified as [15]

$$R_{sum} = \sum_{k} \log_2 \left(1 + \frac{\gamma}{K'} \tilde{r}_{kk}^2\right) \quad \text{bit/s/Hz}, \tag{9}$$

where $\tilde{r}_{kk}$ is the k-th element of $\text{diag}(\mathbf{R}(1 : K', 1 : K'))$.

For the IUs, this stage requires $\text{Card}(\mathcal{K}_{IN})$ iterations and $\text{Card}(\mathcal{G})$ QR decomposition updates in the i-th iteration. Thus, the total complexity is $\mathcal{O}(2\text{Card}(\mathcal{V})\text{Card}(\mathcal{K}_{NIU})^2)(\text{step } 12) + \mathcal{O}((3\text{Card}(\mathcal{G})^2\text{Card}(\mathcal{K}_{IN})^2 + 2\text{Card}(\mathcal{G})\text{Card}(\mathcal{K}_{IN})^3)/3)$.

IV. Performance Evaluation

This section presents the results for the proposed BS algorithms. We have gauged their performance against benchmark algorithms, D-QR-P, I-QR-P, IA-BS [4], MM-BS [5], and maximizing SINR [6], for a fair comparison. Note that, we do not include the results for C-QR-P[2].

[2] It was already proven in [15] that I-QR-P and D-QR-P provide almost similar sum-rate with the C-QR-P but with considerably less complexity.

We generate a mmWave B-MIMO, where the base station is a ULA with N antennas serving K randomly distributed single-antenna users simultaneously. The channel has one LoS path defined by $\alpha_k^{(0)} \sim \mathcal{CN}(0, 1)$ and two NLoS paths given as $\alpha_k^{(1,2)} \sim \mathcal{CN}(0, 10^{-2})$ where the spatial direction φ is uniformly distributed over $[-\pi/2, \pi/2]$. Further, all results are produced on a computer with a 16 GB RAM and 3.4 GHz Intel i7-6700 CPU, and averaged over 500 channel realizations.

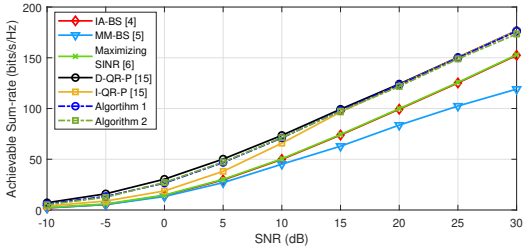

Fig. 2: Achievable sum-rate vs SNR, where $N = 256$ and $K = 16$.

The achievable sum-rate performance of the algorithms is compared in Fig. 2. It is evident from the plot that the proposed algorithms outperform IA-BS, MM-BS, and maximizing SINR while they perform almost identically with I-QR-P and D-QR-P at high SNRs. However, I-QR-P suffers a slight performance loss at low SNRs, which can be compensated by Algorithm 2. This is because I-QR-P starts with the QR decomposition of only one row (i.e., a beam), which provides limited information about the sparse channel, and adds new beams iteratively. In contrast, Algorithm 2 starts with the QR decomposition of the sparse channel matrix acquired for the NIUs, which delivers more information than I-QR-P at the beginning.

The effect of beam diversity is evaluated in Fig. 3a. Since Algorithm 1 and 2 choose M adaptively to provide enough diversity, as mentioned in Fig. 1, they outperform the other fixed cases. The case of $M = 2$ has poor performance, especially for Algorithm 2, due to not having enough diversity.

The increment in beam resolution increases the sum-rate for the proposed algorithms, as shown in Fig. 3b. However, depending on the sparsity, we obtain a performance gap between them. When high sparsity (i.e., $K \ll N$) exists, this performance gap decreases, so that Algorithm-1 is superior to Algorithm 2 in case of low sparsity.

Next, we compare the average running times in Fig. 3c and Fig. 3d. Note that both do not provide the result for D-QR-P since I-QR-P enjoys much less complexity than D-QR-P [15].

In Fig. 3c, both Algorithm 1 and 2 perform considerably faster BS than I-QR-P. Note that Algorithm 2 outperforms others and its speed-up factor gets more prominent as N increases since it decreases the search size significantly in stage 2 and selects beams for the IUs from the set $\mathcal{G}$ in stage 3. A low complexity BS for the NIUs is already performed in stage 1. However, I-QR-P selects beams for all K users from the beam set $\mathcal{D}$, leading to increased complexity. Fig. 2 and Fig. 3c reveal that reducing the size of candidate beams is critical in speeding up the BS while maintaining almost the same sum-rate. In other words, if a beam contributes to no user, there is no need to consider it in the selection process.

The effectiveness of the proposed algorithms for a sparse environment is evaluated in Fig. 3d. In this case, Algorithm 1 and 2 still select beams faster than the benchmark. The average

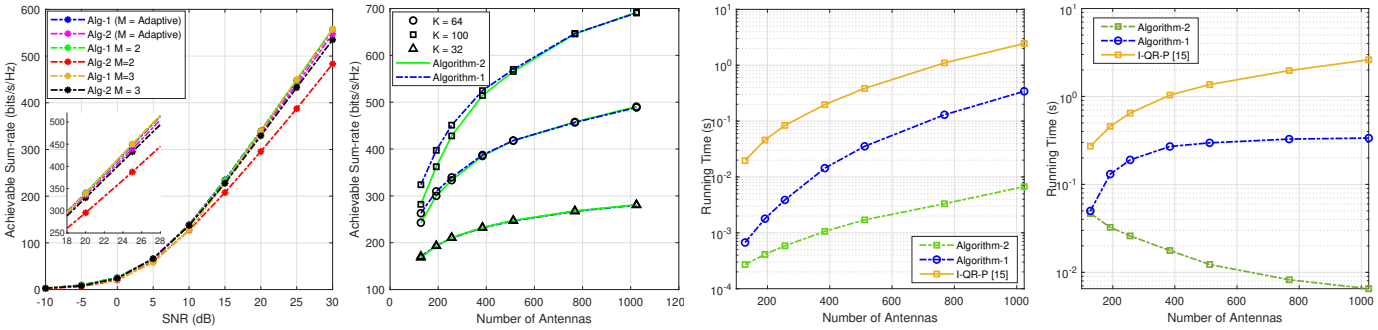

Fig. 3: (a) Effect of M on sum-rate, where $N = 256$ and $K = 64$, (b) Effect of the sparsity on sum-rate, (c) Averaged running times of the algorithms, where $K = N/16$, (d) Averaged running times of the algorithms when $K = 64$ is fixed. (N varies from 128 to 1024 in (b), (c), and (d))

run time reduces for Algorithm 2 as the sparsity increases since the probability of choosing the same strongest beam, $P = 1 - \frac{N!}{N^K(N-K)!}$, in stage 1 decreases as N grows when $K = 64$ is fixed. For example, $P \approx 99\%$ when $N = 256$, $P \approx 93\%$ when $N = 768$, and $P \approx 82\%$ when $N = 1024$. Since the decrease in P and increase in beam diversity occur as N grows, it is likely to have fewer IUs in stage 1. Thus, most users directly select their strongest beams in stage 1 because NIUs outnumber IUs. Consequently, the number of iterations decreases to select beams with QR update for IUs in stage 3. The behavior of Algorithm 2 is different in Fig. 3c and Fig. 3d. This is because the simulation setup in Fig. 3c, where $K = N/16$), causes an increase in P. For example, $P \approx 38\%$ when $N = 256$, $P \approx 78\%$ when $N = 768$, and $P \approx 87\%$ when $N = 1024$. Therefore, there will be more IUs with this setup as N grows, increasing the complexity in stage 3.

V. Conclusion

This letter proposes a two-stage BS algorithm to reduce the complexity of existing D-QR-P and a three-stage BS algorithm for I-QR-P. The benchmark algorithms select beams from all available beams during the BS process, although most do not contribute to users. Nevertheless, the proposed algorithms first identify the most contributing beams to narrow the candidate beam set and perform the BS afterward. Combining this strategy with matrix perturbation theory reduces the BS complexity significantly. The results validate that the proposed algorithms provide almost identical sum-rate performance with the baseline algorithms, and both can be adopted at low and high SNRs. Note that I-QR-P is only suitable for application at medium and high SNRs due to the loss observed in low SNRs. Thence, the three-stage method is more attractive since it can compensate for this loss along with its lowest complexity.

Additionally, the practical implementation of proposed methods for wideband mmWave scenarios will be investigated with multi-antenna users as a future study.

Acknowledgments

This work was supported in part by the U.S. National Science Foundation under Grant ECS-1923857.

References

[1] M. M. Lodro, N. Majeed, A. A. Khuwaja, A. H. Sodhro, and S. Greedy, "Statistical channel modelling of 5G mmWave MIMO wireless communication," in *Proc. Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, 2018, pp. 1–5.

[2] J. Huang, C.-X. Wang, R. Feng, J. Sun, W. Zhang, and Y. Yang, "Multi-frequency mmWave massive MIMO channel measurements and characterization for 5G wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1591–1605, 2017.

[3] D. J. Love and R. W. Heath, "Multimode precoding for MIMO wireless systems," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3674–3687, 2005.

[4] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, 2016.

[5] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, 2013.

[6] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, 2015.

[7] I. Orikumhi, J. Kang, H. Jwa, J.-H. Na, and S. Kim, "SINR maximization beam selection for mmWave beamspace MIMO systems," *IEEE Access*, vol. 8, pp. 185 688–185 697, 2020.

[8] D. Wang, W. Zhang, and Q. Zhu, "Heuristic search inspired beam selection algorithms for mmWave MU-MIMO system with discrete lens array," *IEEE Access*, vol. 9, pp. 61 324–61 333, 2021.

[9] M. A. L. Sarker, M. F. Kader, and D. S. Han, "Rate-loss mitigation for a millimeter-wave beamspace MIMO lens antenna array system using a hybrid beam selection scheme," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3582–3585, 2020.

[10] S. Cetinkaya, L. Afeef, G. Mumcu, and H. Arslan, "Heuristic inspired precoding for millimeter-wave MIMO systems with lens antenna subarrays," in *Proc IEEE 95th Veh. Technol. Conf. (VTC2022-Spring)*, 2022, pp. 1–6.

[11] W. Shen, X. Bu, X. Gao, C. Xing, and L. Hanzo, "Beamspace precoding and beam selection for wideband millimeter-wave MIMO relying on lens antenna arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 24, pp. 6301–6313, 2019.

[12] C. Feng, W. Shen, and J. An, "Beam selection for wideband millimeter wave MIMO relying on lens antenna arrays," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1875–1878, 2019.

[13] R. Pal, A. K. Chaitanya, and K. V. Srinivas, "Low-complexity beam selection algorithms for millimeter wave beamspace MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 768–771, 2019.

[14] R. Pal, K. V. Srinivas, and A. K. Chaitanya, "A beam selection algorithm for millimeter-wave multi-user MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 852–855, 2018.

[15] Q. Zhang, X. Li, B.-Y. Wu, L. Cheng, and Y. Gao, "On the complexity reduction of beam selection algorithms for beamspace MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1439–1443, 2021.

[16] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.

[17] M. Nazzal, M. A. Aygül, A. Görçin, and H. Arslan, "Dictionary learning-based beamspace channel estimation in millimeter-wave massive MIMO systems with a lens antenna array," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, 2019, pp. 20–25.

[18] G. H. Golub and C. F. V. Loan, *Matrix Computations, 3rd Ed.* Baltimore, MD, USA: The John Hopkins University Press, 1996.