Chapter 3

Beam Selection Algorithms for mmWave beamspace MU-MIMO System

3.1 Introduction

One main issue with scaling up the number of beams in mmWave beamspace MU-MIMO systems relates to the cost and complexity of RF components which are deployed with antenna elements. To circumvent this problem, beam selection techniques are used and the number of required RF chains are reduced while keeping most of the advantages of mmWave beamspace MU-MIMO systems. The principle idea of beam selection scheme is to connect a limited number of RF chains to a subset of beams optimally among all available beams in transmitter. However, mmWave beamspace MU-MIMO systems with more beams lead to increase the complexity of the system due to the high number of computations required during optimal beam selection. Many beam selection algorithms have been introduced to alleviate the computation process without considerable loss in the system performance [108–110].

Recently, a lot of interest has emerged to analyze the performance of beams selection algorithm for mmWave beamspace MU-MIMO systems. The problem of the maximizing the magnitude (MM), maximizing the signal-to-interference-plus-noise-ration (M-SINR), and maximizing the capacity (MC) with a few beams has been considered and has been solved iteratively [108, 109]. These methods exhibit a good system performance with few number of beams but it tends to be more challenging when the number of beams increase. A suboptimal algorithm of beam selection based on the “MM”, “M-SINR”, and “MC” is proposed in [108, 109]. Since optimum method of beam selection requires exhaustive search over all possible combinations of beam subsets, it is impractical to use when there are a large number of beams. Any beam selection algorithm should be designed to reduce the complexity of beam selection selection process as well as achieve the required level of spectral efficiency. Motivated by the idea of selecting beams depending on the sum of logarithms of square of Eigenvalue of their effective channels, this chapter propose a beam selection algorithm for mmWave beamspace MU-MIMO downlink systems. At each iteration, the exclusive beam which maximizes the sum of logarithms of square of Eigenvalue of their effective channels obtained by a set of beams is selected and deactivated. The Eigenvalues of each user's effective channel can be obtained from that of the QR decomposition of the beamspace channel matrix. Hence, refer to it as "QR-based" beam selection. The selection process repeats until the algorithm reach the required subset of beams. The iterative precoding method proposed in [53] is applied to cancel the MUI in the precoding stage of the beam selection process. The precoding algorithm applies the QR decomposition to reduce the computational burden through beam selection process.

At the same time, it should be noted that the transmitter can be equipped with a very large number of antenna elements at mmWave frequencies, making N to be very high. It means that the number of available beams will be large, and it has been observed that the cell radius in mmWave cellular systems is small which is order of a few meters, i.e., 50 to 100 meters. Hence, the number of active users per cell would be relatively low, i.e., $N \gg K$. This entails that the mmWave cellular systems are likely to encounter sparse systems more than dense systems. In this context, a greedy and MWM-based beam selection algorithm are proposed for sparse system. Greedy beam selection is a search based iterative procedure, which allocates the strongest beam to the corresponding user. If a beam is the strongest for more than one user then that beam will be allocated to the user having a higher channel gain. For the remaining users, algorithm repeats the process after removing the allocated beams from the set of available beams. This process is repeated until a beam is selected for each user, whereas beam selection through MWM algorithm is to find a matching beam that maximizes the sum of edge weights over the bipartite graph. The edge weights are nothing but SINRs achieved by each user through each beam, and well efficient Kuhn-Munkres algorithm [111, 112] solves the MWM problem efficiently.

3.2 mmWave Beamspace Downlink MU-MIMO Communication System Model

We are considering a downlink mmWave beamspace MU-MIMO communication system having a transmitter equipped with N beams, and K users equipped with a single receive antenna. Thus, the input-output relation for a mmWave beamspace MU-MIMO system [113] can be expressed as

$$\mathbf{y}_b = \mathbf{H}_b^H \mathbf{P}_b \mathbf{x} + \mathbf{w}_b, \tag{3.1}$$

where $\mathbf{H}_b^H = \mathbf{H}^H \mathbf{U} = [\mathbf{h}_{b,1}^H, \ldots, \mathbf{h}_{b,K}^H]^H \in \mathbb{C}^{K \times N}$ is the beamspace channel matrix. Each $\mathbf{h}_{b,k}^H = \mathbf{h}_k^H \mathbf{U} \in \mathbb{C}^{N \times 1}$, $k = 1, \ldots, K$, and $\mathbf{H}^H$ is the spatial channel matrix. $\mathbf{P}_b \in \mathbb{C}^{N \times K}$ is a digital precoder to remove MUI. $\mathbf{x} \in \mathbb{C}^{K \times 1}$ is the transmitted symbol vector while satisfying an average power constraint as $\mathbb{E}[\| \mathbf{P}_b \mathbf{x} \|^2] \leq \rho$, where ρ is the total transmitted power. $\mathbf{y}_b \in \mathbb{C}^{K \times 1}$ is the received information vector, and $\mathbf{w}_b \in \mathbb{C}^{K \times K}$ denotes AWGN noise vector with $\mathbf{w}_b \sim \mathcal{CN}(0, N_0 \mathbf{I}_K)$.

Further, the input-output relation for a mmWave beamspace MU-MIMO system, after beam selection, can be expressed as

$$\tilde{\mathbf{y}}_b = \tilde{\mathbf{H}}_b^H \tilde{\mathbf{P}}_b \mathbf{x} + \tilde{\mathbf{w}}_b, \tag{3.2}$$

where $\tilde{\mathbf{H}}_b^H = [\tilde{\mathbf{h}}_{b,1}^H, \ldots, \tilde{\mathbf{h}}_{b,K}^H]^H \in \mathbb{C}^{K \times K}$ is the beamspace channel matrix, and the columns are corresponding to the K selected beams. $\tilde{\mathbf{P}}_b \in \mathbb{C}^{K \times K}$ is a digital precoding matrix. $\tilde{\mathbf{w}}_b$ denotes AWGN noise vector with $\tilde{\mathbf{w}}_b \sim \mathbb{C}\mathcal{N}(0, N_0 \mathbf{I}_K)$.

3.3 Proposed Beam Selection Algorithm and Precoding Scheme

In the following section, a “QR-based” beam selection algorithm is proposed with pre –cancelling the interference. Further, the proposed iterative precoding scheme to pre –cancel interference while performing the “QR-based” beam selection is discussed. The details of this algorithms are given below.

3.3.1 QR-based Beam Selection

"QR-based" beam selection algorithm works iteratively to select beams, and it selects a beam that contributes least to the sum rate of users' effective channels and deactivates it. In other words, the algorithm maximizes the sum of logarithms of square of Eigenvalue of their effective channels with deleting a beam, is selected and deactivates in each iteration. In descending order, the algorithm repeats until the required number of beams are obtained. Further, the proposed algorithm uses the iterative precoding design to pre-cancel MUI. Let the QR decomposition of $\tilde{\mathbf{H}}_b$ is given by

$$\tilde{\mathbf{H}}_b^H = (\tilde{\mathbf{Q}}_b \tilde{\mathbf{R}}_b)^H, \tag{3.3}$$

where $\tilde{\mathbf{Q}}_b = [\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_K] \in \mathbb{C}^{K \times K}$ is a unitary matrix and $\tilde{\mathbf{R}}_b \in \mathbb{C}^{K \times K}$ is an upper triangular matrix [114]. The columns of $\tilde{\mathbf{Q}}_b$ provide an orthonormal basis, obtained through Gram-Schmidt procedure (for more details, refer to Appendix A), for the vector space spanned by the K column vectors of $\tilde{\mathbf{H}}_b$. By choosing $\tilde{\mathbf{P}}_b = \tilde{\mathbf{Q}}_b,$ (3.2) becomes

$$\tilde{\mathbf{y}}_b = \tilde{\mathbf{R}}_b^H \mathbf{x} + \tilde{\mathbf{w}}_b, \tag{3.4}$$

where $\tilde{\mathbf{R}}_b^H$ can be expressed as

$$\tilde{\mathbf{R}}_b^H = \begin{vmatrix} \tilde{r}_{b,11} & 0 & \dots & 0 \\ \tilde{r}_{b,21} & \tilde{r}_{b,22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{r}_{b,K1} & \tilde{r}_{b,K2} & \dots & \tilde{r}_{b,KK} \end{vmatrix} \tag{3.5}$$

Thus, when K data streams are multiplexed in the coordinate system specified by $\tilde{\mathbf{Q}}_b$, the received signal at user k becomes

$$\tilde{y}_{b,k} = \tilde{r}_{b,kk} x_k + I_{b,k} + \tilde{w}_{b,k},\ k \in \{1, \dots, K\}, \tag{3.6}$$

where $I_{b,k} = \sum_{k>j} \tilde{r}_{b,kj} x_j$ is the interference signal. Interference can be made equal to zero for all the users by diagonalizing $\tilde{\mathbf{R}}_b^H$. We achieve this by having $\tilde{\mathbf{P}}_b = \tilde{\mathbf{Q}}_b \mathbf{L}$ in (3.2), where one can obtain,

$$
\begin{aligned}
\tilde{\mathbf{y}}_b &= \tilde{\mathbf{H}}_b^H \tilde{\mathbf{P}}_b \mathbf{x} + \mathbf{w}_b, \\
\tilde{\mathbf{y}}_b &= \tilde{\mathbf{R}}_b{}^H \mathbf{L} \mathbf{x} + \mathbf{w}_b.
\end{aligned} \tag{3.7}
$$

We compute (in the sub-section 3.3.3) $\mathbf{L} \in \mathbb{C}^{K \times K}$ such that $\tilde{\mathbf{R}}_b^H \mathbf{G}$ is a diagonal matrix, with its K^{th} diagonal entry equal to $\tilde{r}_{b,kk}$. With $I_{b,k} = 0$, R_k depends only on $\tilde{r}_{b,kk}$, the effective channel gain for user k, and the sum-rate is given by

$$
\begin{aligned}
R_s &= \sum_k R_k \ \text{ bits/s/Hz} \\
R_s &= \sum_k \log_2 \left(1 + \frac{1}{N_0} \frac{\rho}{K} \tilde{r}_{b,kk}^2 \right) \ \text{ bits/s/Hz}
\end{aligned} \tag{3.8}
$$

where N_0 is the noise power of the AWGN at each user.

3.3.2 Algorithms for Computing $\tilde{\mathbf{H}}_b$

The algorithm consists of $(N - K)$ iterations. In each iteration, the algorithm eliminate a beam (i.e., a row of $\mathbf{H}_b$) that contributes minimally towards R_s. To make a decision on which row to be removed, the algorithm follow another iterative process. Reduced-dimensional channel matrix at the end of $(i-1)^{\text{th}}$ iteration, denoted by $\mathbf{H}_b^{(i)}$, can be expressed as

$$
\mathbf{H}_b^{(i)} = \left[\left(\mathbf{c}_1^{(i)} \right)^T \ \left(\mathbf{c}_2^{(i)} \right)^T \ \ldots \ \left(\mathbf{c}_{N-i}^{(i)} \right)^T \right]^T, \tag{3.9}
$$

where $\mathbf{c}_j^{(i)} \in \mathbb{C}^{1 \times K}$ is the j^{th} row of $\mathbf{H}_b^{(i)}$, for $j = 1, \ldots, N-i$. From $\mathbf{H}_b^{(i)}$, the algorithm delete one row (out of $N-i$ rows), by following another iterative process, which can be explained as follows.

Let $\mathbf{T}^{(i)} = \mathbf{H}_b^{(i)}$. For $j = 1, \ldots, N-i$, the algorithm remove $\mathbf{c}_j^{(i)}$ from $\mathbf{T}^{(i)}$ to obtain the matrix $\mathbf{T}_{-j}^{(i)}$, and find its QR decomposition

$$
\mathbf{T}_{-j}^{(i)} = \mathbf{Q}_{-j}^{(i)} \mathbf{R}_{-j}^{(i)}. \tag{3.10}
$$

Next, compute

$$\gamma_j^{(i)} = \sum_{u \in \{1,\ldots,N-i\}} \log_2 \left(1 + \left(r_{-j_{uu}}^{(i)} \right)^2 \right), \tag{3.11}$$

where $r_{-j_{uv}}^{(i)}$ denotes the $(u, v)^{\text{th}}$ element of $\mathbf{R}_{-j}^{(i)}$. Please note that the interference terms have been ignored while computing $\gamma_j^{(i)}$. This will not be a problem as precoder $\tilde{\mathbf{P}}_b$ will completely cancel the MUI. The algorithm remove row j' from $\mathbf{H}_b^{(i)}$ to obtain $\mathbf{H}_b^{(i+1)}$, where

$$j' = \underset{j \in \{1,\ldots,N-i\}}{\arg\max} \ \gamma_j^{(i)}. \tag{3.12}$$

Thus, the algorithm eliminate a beam whose effect on the sum-rate is the least. Please refer to Algorithm 1 for more details.

Algorithm 1 Algorithm to compute $\tilde{\mathbf{H}}_b$

Initialize $\mathbf{H}_b^{(0)} = \mathbf{H}_b$

for $i = 0 \to N - K - 1$ **do**

  $\mathbf{T}^{(i)} = \mathbf{H}_b^{(i)}$

  for $j = 1 \to N - i$ **do**

    Remove row j from $\mathbf{T}^{(i)}$ to obtain $\mathbf{T}_{-j}^{(i)}$

    $\mathbf{T}_{-j}^{(i)} = \mathbf{Q}_{-j}^{(i)} \mathbf{R}_{-j}^{(i)}$ (QR-decomposition of $\mathbf{T}_{-j}^{(i)}$)

    $\gamma_j^{(i)} = \sum_{u=1}^{N-i} \log_2 \left(1 + \left(r_{-j_{uu}}^{(i)} \right)^2 \right)$

  end for

  $j' = \arg\max_{j \in \{1,\ldots,N-i\}} \gamma_j^{(i)}$

  Remove $\mathbf{c}_{j'}^{(i)}$ from $\mathbf{H}_b^{(i)}$ to obtain $\mathbf{H}_b^{(i+1)}$

end for

$\tilde{\mathbf{H}}_b = \mathbf{H}_b^{(N-K)}$

3.3.3 **Proposed Precoding Scheme**

Now, using $\tilde{\mathbf{H}}_b$ given by the Algorithm 1 and by choosing $\tilde{\mathbf{P}}_b = \tilde{\mathbf{Q}}_b$, (with $\tilde{\mathbf{Q}}_b$ obtained from the QR decomposition of $\tilde{\mathbf{H}}_b = \tilde{\mathbf{Q}}_b \tilde{\mathbf{R}}_b$), one can obtain the system described by (3.6). Interference $I_{b,k}, k = 1, \ldots, K$, can be eliminated by diagonalizing $\tilde{\mathbf{R}}_b^H$. This section propose to accomplish it by having $\tilde{\mathbf{P}}_b = \tilde{\mathbf{Q}}_b \mathbf{L}$ and by computing $\mathbf{L} \in \mathbb{C}^{K \times K}$ such that $\tilde{\mathbf{R}}_b^H \mathbf{L}$ is a diagonal matrix, with its k^{th} diagonal entry is given by $\tilde{r}_{b,kk}$.

An element $\tilde{r}_{b,kj},\ k > j$, in $\tilde{\mathbf{R}}_b^H$ can be made equal to zero by post-multiplying $\tilde{\mathbf{R}}_b^H$ with a matrix. Suppose $\mathbf{R}' = \tilde{\mathbf{R}}_b^H \mathbf{E}^{(k,j)}$, where $\mathbf{E}^{(k,j)}$ is a $K \times K$ lower triangular matrix obtained as follows: for $l = 1, \ldots, K$, $m = 1, \ldots, K$,

$$e_{lm}^{(k,j)} = \begin{cases} 1, & \text{if } l = m, \\ -\frac{\tilde{r}_{b,kj}}{\tilde{r}_{b,kk}}, & \text{if } l = k, m = j, \\ 0, & \text{otherwise.} \end{cases} \tag{3.13}$$

$\mathbf{L}$ is computed as a product of elementary matrices defined in (3.13) taken in a specific order. Algorithm 2 details how to compute $\mathbf{L}$. We observe that the matrix $\mathbf{L}$ is a lower triangular matrix. Please note that, if $\tilde{r}_{b,kk} \neq 0$ for all $k = 1, 2, \ldots, K$, the matrix $\mathbf{L}$ computed using Algorithm 2 will diagonalize $\tilde{\mathbf{R}}_b^H$.

Algorithm 2 Algorithm to compute $\mathbf{L}$

Initialize $\mathbf{L} = \mathbf{I}_K$, where $\mathbf{I}_K$ is identity matrix of order K

Initialize $\mathbf{R}' = \tilde{\mathbf{R}}_b^H$

for $k = 1 \to K$ **do**

  if $r'_{kk} \neq 0$ **then**

    for $j = 1 \to k-1$ **do**

      Define $\mathbf{E}^{(k,j)}$ as in (3.13)

      Compute $\mathbf{R}' = \mathbf{R}'\mathbf{E}^{(i,j)}$

      $\mathbf{L} = \mathbf{L}\mathbf{E}^{(k,j)}$

    end for

  end if

end for

$\tilde{\mathbf{R}}_b^H$ is not diagonalizable, if $\tilde{r}_{b,kk} = 0$ for some $k \in \{1, \ldots, K\}$. In such a case, if $\tilde{r}_{b,kk} = 0$ for some $k \in \{1, 2, \ldots, K\}$, then $\tilde{y}_k$ does not contain the message x_k intended for user k and the particular user need not be served. By considering the effective channel matrix for the remaining $K-1$ users, we can still use Algorithm 2 to compute $\mathbf{L}$ that avoids interference at all users with $\tilde{r}_{b,kk} \neq 0$. In addition, when the magnitude of any particular $\tilde{r}_{b,kk}$ is significantly low compared to other diagonal values of $\tilde{\mathbf{R}}_b^H$, making $\tilde{\mathbf{R}}_b^H$ ill-conditioned, and serve the other users. Hence, there is no need to avoid interference at user k as it can not decode its message even in the absence of interference. Thus, even if $\tilde{\mathbf{R}}^H$ is not diagonalizable, Algorithm 2 can be used to compute $\mathbf{L}$ that avoids interference at all users with $\tilde{r}_{kk} \neq 0$.

Note that $\mathbf{L}$ need not be unitary and to satisfy the transmit power constraint, one needs to normalize $\mathbf{L}$ to have unit norm. Thus, $\tilde{\mathbf{P}}_b$ in (3.2) is chosen as $\tilde{\mathbf{P}}_b = \tilde{\mathbf{Q}}_b\tilde{\mathbf{L}}$, where $\tilde{\mathbf{L}} = \frac{\mathbf{L}}{\|\mathbf{L}\|_F}$, $\tilde{\mathbf{Q}}_b$ is obtained from the QR decomposition of $\tilde{\mathbf{H}}_b = \tilde{\mathbf{Q}}_b\tilde{\mathbf{R}}_b$.

3.3.4 Complexity Analysis

The main complexity of the above algorithm is due to the QR decomposition in the inner loop. In i^{th} iteration, $i = 0, \ldots, (N-K-1)$, the algorithm computes $(N-i)$ QR decomposition and the complexity of each decomposition is $\mathcal{O}(2(N-i)K^2)$ [114]. Thus, the overall complexity is $\sum_{i=0}^{N-K-1} (N-i)\mathcal{O}(2(N-i)K^2)$. Similarly, complexity of the "IA" beam selection algorithm is $\sum_{i=0}^{K-NIU-1} (N-NIU-i)\mathcal{O}(K^3)$, where NIU is the number of non-interfering users, and complexity of "MC" beam selection algorithm is $\sum_{i=0}^{N-K-1} (N-i)\mathcal{O}(2(N-i)K^2+K^3+(N-i)^2K)$. Eventually, this complexity is higher than that of the "IA" beam selection [110] and lower than "MC" beam selection [109].

3.3.5 Numerical Results

In this section, we evaluate the performance of the proposed beam selection algorithm through simulations, and compare it with the existing beam selection algorithms. We focus on two performance metrics, namely, spectral efficiency and power efficiency.

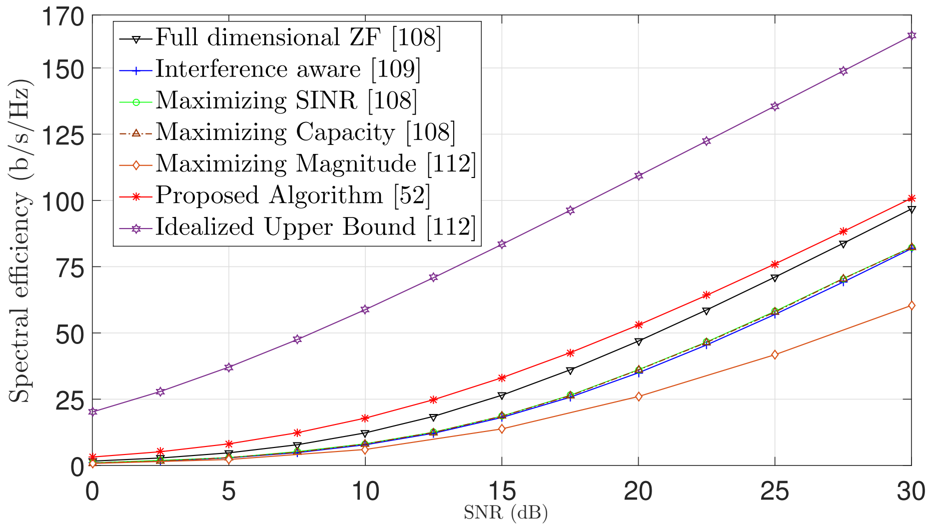

Figure 3.1 Sum-rate Performance Comparison

We consider downlink mmWave beamspace MU-MIMO communication system having a transmitter equipped with $N = 256$ beams and $K = 16$ users, each user having a single receive antenna. Spatial channel between the AP and user $k, k \in \{1, \dots, K\}$, is assumed to be having one LoS component with path gain $\beta_k^{(0)} \sim \mathcal{CN}(0, 1)$ and two NLoS components, each having a path gain $\beta_k^{(\ell)} \sim \mathcal{CN}(0, 10^{-2}), \ell = 1, 2$. The path gains $\beta_k^{(\ell)}$ are assumed to be independent of each other. The spatial frequencies, $\theta_k^{(\ell)}, k = 0, 1, 2$, of user k, are uniformly distributed in the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$ and independent of each other. Fig. 3.1 plots the achievable sum-rate of the proposed algorithm and the performance is compared with "MM" beam selection with one beam per user [113], [109], "MC", "M-SINR" [109], and "IA" beam selection [110]. There are two types of "MC" algorithms – an "MC-incremental" and an "MC"- decremental. Performance of both these algorithms is almost the same, and we have compared the results with the "MC-decremental" algorithm. It can be observed that the proposed algorithm outperforms the other algorithms. In "MC" beam selection, the beams are selected by maximizing the capacity and it is given by (24) in Section 3C in [109] as

$$R(\mathbf{H}_b) = K \log \det \left(1 + \frac{\rho}{N_0} \mathbf{H}_b^H \mathbf{H}_b\right) \text{ bits/s/Hz.}$$

It should be noted that (24) is the maximum achievable rate of a point-to-point MIMO channel and maximizing (24) may not maximize the sum-rate of a mmWave

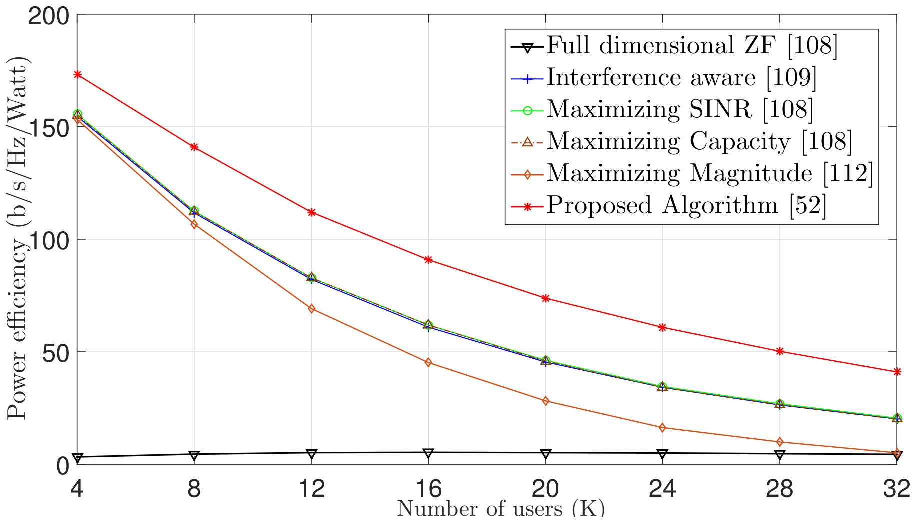

Figure 3.2 Power Efficiency Comparison

beamspace MU-MIMO system that is under consideration.

"Full dimensional ZF" is the sum-rate achieved by full system ZF precoding that uses all the N beams. Note that, full dimensional ZF is one of the upper bounds on the achievable performance of a full dimensional system. Full dimensional Wiener filtering, full dimensional matched filtering and the idealistic upper bound, proposed in [113], provide us with other upper bounds on the full dimensional system performance. The idealistic upper bound is given by (13) in [113] as

$$R_{up} = K \log \left(1 + \rho \frac{N}{K}\right) \text{ bits/s/Hz.}$$

We note that the sum-rate achieved by the proposed beam selection algorithm is considerably lower than the idealistic upper bound, but higher than the performance of full dimensional ZF.

Fig. 3.2 compares the power efficiency of different beam selection algorithms, including the proposed one. Power efficiency is defined as [109, 110, 115],

$$\eta_p = \frac{R_s}{\rho + N_{\text{RF}} P_{\text{RF}}}, \text{ bits/s/Hz/Watt,} \tag{3.14}$$

where N_{RF} is the number of RF chains and P_{RF} is the power consumed by an RF chain. As in [109, 110], we evaluated η_p at SNR $= 20$ dB, with $P_{\text{RF}} = 34.4$ mW and

$\rho = 32$ mW. As it is evident from the figure that the proposed method achieves considerably higher power efficiency than the other beam selection algorithms. Note that the full dimensional ZF has very high power consumption. It is worth noting that different users see different effective channel gains and optimal power allocation across users might further improve the sum-rate with the proposed algorithm.

3.4 Proposed Low Complexity Beam Selection Algorithms

In this section, low complexity beam selection — greedy and MWM-based — algorithm is proposed. The details of these two beam selection algorithms are given below.

3.4.1 Greedy Beam Selection

This section now proposes a greedy algorithm that allocates a beam to each user. Let $\mathbf{G} \in \mathbb{R}^{K \times N}$ be the channel gain matrix, where its $(k,n)^{\text{th}}$ element g_{kn} is given by the squared absolute value of the $(k,n)^{\text{th}}$ element of $\mathbf{H}_{\text{b}}^{H}$. Beam n is considered stronger to user k_1 than user k_2, if $g_{k_1 n} > g_{k_2 n}$. Greedy beam selection, presented in Algorithm 3, is a search based iterative procedure. Let β_k be the index of the strongest beam for user k. In the first iteration, one can find $\beta_k = \arg\max_{n \in \{1,\dots,N\}} g_{kn},\ k = 1,\dots,K$. If $\beta_k \neq \beta_{k'}$ for $k' \in \{1,\dots,K\} \setminus \{k\}$, i.e., if beam β_k is the strongest beam only for user k, then β_k is allocated to user k. If a beam is the strongest beam for more than one user, then that beam will be allocated to the user having higher channel gain. In other words, if $\beta_k = \beta_{k'}$ for $k \neq k'$, then β_k is allocated to k if $g_{kn} > g_{k'n}$; otherwise it is allocated to k'. For the remaining users, the process is repeated after removing the allocated beams from the set of available beams. This process is repeated until a beam is selected for each user. Note that Algorithm 3 allocates a beam to each user provided $N \geq K$.

Number of iterations in the greedy algorithm depends on the probability that a given beam is the strongest to more than one user. Next, the probability of two or more users sharing the strongest beam is computed and it is shown that Algorithm 3 performs beam selection in less number of iterations if $K \ll N$. Let the channel gains g_{kn} be identical and independently distributed and let $F_g(x)$ and $f_g(x)$ be the cumulative distribution function and probability density function, respectively, of $g_{kn}, \forall k, n$. Without loss of generality, it is considered that a user k' and a beam n' and compute the probability, $p_{k'n'}$, that n' is the strongest beam for k'.

$$\begin{aligned} p_{k'n'} =& \Pr\{g_{k'n'} > g_{k'n}, \forall n \neq n'\} \\ =& \int_0^\infty \left(\Pr\{g_{k'n} < x, \text{for some } n\}\right)^{N-1} dF_g(x), \\ =& \frac{1}{N}. \end{aligned} \tag{3.15}$$

Thus, $\frac{1}{N}$ is the probability that a given beam is the strongest to a given user.

The probability p that a given beam, out of N, is the strongest to two or more users is given by

$$\begin{aligned} p =& \sum_{i=2}^{K} \binom{K}{i} \left(\frac{1}{N}\right)^i \left(1 - \frac{1}{N}\right)^{K-i} \\ =& 1 - \left(1 - \frac{1}{N}\right)^K - \frac{K}{N}\left(1 - \frac{1}{N}\right)^{K-1}. \end{aligned} \tag{3.16}$$

For a fixed value of K, $p \to 0$ as $N \to \infty$. Thus, if N is large compared to K, the probability of two or more users having the same beam as their strongest beam is small. One can call such a system, a sparse system. $p \to 1$ if $N \to \infty$ with $\frac{K}{N} \approx 1$; i.e., if the values of K and N are close to each other, with high probability, there exists a beam that is stronger to more than one user. This is referred to such a system as a dense system. The proposed greedy algorithm requires less number of iterations in a sparse system as compared to the dense system.

3.4.2 MWM-based Beam Selection

This section model the beam selection problem, of selecting K beams out of N, as a MWM problem over a bipartite graph. Let $B = \{b_1, \ldots, b_N\}$ be the set of beams and let $U = \{u_1, \ldots, u_K\}$ be the set of users. Consider a bipartite graph $Z = (V, E)$ where $V = B \cup U$ and $(b_n, u_k) \in E$ for $1 \leq n \leq N$, $1 \leq k \leq K$. Let

Algorithm 3 Greedy Algorithm for Beam Selection

Input: $\mathbf{G}_{K \times N}$
Initialize: $\mathcal{N} = \{1, \dots, N\}$, $\mathcal{K} = \{1, \dots, K\}$
for $k = 1 \to K$ **do**
  $\beta_k = \arg\max_{n \in \{1,\dots,N\}} \{g_{kn}\}$
end for
ITER:
for $k = 1 \to K$ **do**
  if $k \in \mathcal{K}$ **then**
    if $\beta_k \neq \beta_{k'}$ for $k' \in \mathcal{K} \setminus \{k\}$ **then**
      Allocate beam β_k to user k
      $\mathcal{K} \leftarrow \mathcal{K} \setminus \{k\}$ and $\mathcal{N} \leftarrow \mathcal{N} \setminus \{\beta_k\}$
    else
      Allocate β_k to k where, $k = \arg\max_{i \in \mathcal{K}} \{g_{i\beta_k}\}$
      $\mathcal{K} \leftarrow \mathcal{K} \setminus \{\kappa\}$ and $\mathcal{N} \leftarrow \mathcal{N} \setminus \{\beta_k\}$
    end if
  end if
end for
if $\mathcal{K} \neq \emptyset$ **then**
  $\beta_k = \arg\max_{n \in \mathcal{N}} \{g_{kn}\},\ k \in \mathcal{K}$
  Goto ITER
else
  Stop
end if

$w_{nk} = \log_2(1 + \text{SINR}_{nk})$ be the weight assigned to the edge (b_n, u_k), where

$$\text{SINR}_{nk} = \frac{S_{nk}}{I_k + \sigma^2}. \tag{3.17}$$

Here, S_{nk} is the signal power at user k due to beam n, I_k is the interference power at user k from all the beams except beam n, and N_0 is the noise power.

A matching over the bipartite graph Z is defined as a set M, where $M \subseteq E$, which satisfies the following:

1. for each $k \in \{1, \ldots, K\}$, $\exists n \in \{1, \ldots, N\}$ such that $(b_n, u_k) \in M$, i.e., each user will get at least one beam,

2. for each $n \in \{1, \ldots, N\}$, there exists at the most one $k \in \{1, \ldots, K\}$ such that $(b_n, u_k) \in M$, i.e., a beam is allocated to at the most one user.

The problem of MWM is to find a matching M^* that maximizes the sum of edge weights in the matching, which can also be stated as the optimization problem as given below

$$\begin{aligned} \text{maximize} \quad & \sum_{k,n} w_{nk} l_{nk} \\ \text{subject to} \quad & \sum_{n} l_{nk} = 1, k = 1, \ldots, K, \\ & \sum_{k} l_{nk} \leq 1, n = 1, \ldots, N, \\ & l_{nk} \in \{0, 1\}. \end{aligned}$$

MWM is a well studied problem in graph theory. Kuhn-Munkres algorithm [111, 112], solves the MWM problem in an efficient way for $N = K$. Here, the modified Kuhn-Munkres algorithm is employed for the general case of $N \neq K$ in [116] to find MWM over the bipartite graph G, as shown in Fig. 3.3. Kuhn-Munkres algorithm has a complexity of $\mathcal{O}(NK^2)$ [116], whereas the greedy algorithm is a much simpler solution having very low complexity. However, as discussed in previous section, greedy algorithm requires more iterations when the values of N and K becomes comparable to each other. As can be observed from the simulation results presented in section 3.4.6, the MWM framework for beam selection does not have any such limitations and achieves a better sum rate, even for dense systems.

While computing SINRs using (3.17), it is considered that all the N beams are active. This results in over estimation of the interference as only K beams are activated at any given time to serve K users. More importantly, interference seen by a user is determined by the beams assigned to the remaining $K1$ users, which implies that, when a user is assigned a beam (i.e., when an edge is selected), all the remaining edge weights change. When the MWM finds a matching, the SINRs will change and it is possible, with the modified edge weights, that a new set of edges maximize the sum rate. This inter-dependency among the edge weights makes the

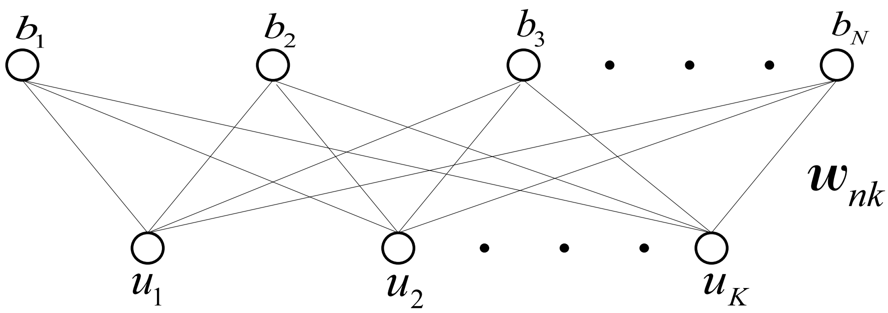

Figure 3.3 Beam selection as MWM over a bipartite graph

MWM based beam selection, presented here, suboptimal. At the same time, this method achieves comparable performance, at significantly low complexity, relative to the existing beam selection algorithms, as discussed in the next section.

3.4.3 **Precoding Scheme**

While beam selection optimally choose beams for the users, precoding, based on the effective channel matrix after beam selection, $\tilde{\mathbf{H}}_{\mathrm{b}}$, is employed to eliminate the interference among the users. ZF is a popular precoding technique. However, ZF degrades the performance when $\tilde{\mathbf{H}}_{\mathrm{b}}$ is ill-conditioned and it can not be applied when $\tilde{\mathbf{H}}_{\mathrm{b}}$ is rank deficient. A precoder that diagonalizes $\tilde{\mathbf{H}}_{\mathrm{b}}$, based on QR decomposition of $\tilde{\mathbf{H}}_{\mathrm{b}}$, has been proposed in section 3.3.3, [53]. We refer to this precoder as "QR-Pr" and employ the same for canceling the interference. After employing the QR-Pr precoder, the received signal at the k^{th} user is given by

$$\tilde{y}_{b,k} = \tilde{r}_{b,kk} x_k + \tilde{w}_{b,k}, \;\; k = 1, \ldots, K, \tag{3.18}$$

where $\tilde{r}_{b,kk}$ is the k^{th} diagonal element of $\tilde{\mathbf{R}}_b$, where $\tilde{\mathbf{H}}_{\mathrm{b}} = \tilde{\mathbf{Q}}_b \tilde{\mathbf{R}}_b$. $\tilde{w}_{b,k}$ is the AWGN with $\sim \mathcal{CN}(0, N_0)$.

3.4.4 **Power Allocation Methodology**

As discussed above, QR-Pr creates a set of K parallel channels, with $\tilde{r}_{kk}$ as the gain of k^{th} channel. Assuming Gaussian signalling, it is well known that water-filling maximizes the sum rate across parallel channels [40]. Thus, instead of uniform power allocation, we propose to employ water-filling power allocation which is given by

$$\begin{aligned} \text{maximzes} \quad & \textstyle\sum_{k=1}^{K} \log_2 \left(1 + \frac{\tilde{r}_{b,kk}^2 p_k}{N_0}\right), \\ \text{subject to} \quad & \textstyle\sum_{k=1}^{K} p_k \le \rho, \\ & p_k \ge 0, \forall k. \end{aligned} \tag{3.19}$$

where p_k is the power allocated to user k, and ρ is the total transmitted power. The optimization problem (3.19) is a classical power allocation problem and its solution is water-filling power allocation and is given by

$$\begin{aligned} p_k &= \left\{\left(\lambda - \frac{N_0}{\tilde{r}_{b,kk}^2}\right)\right\}^+, \quad k = 1, \dots, K, \\ p_k &= \max\left\{0, \left(\lambda - \frac{N_0}{\tilde{r}_{b,kk}^2}\right)\right\}, \quad k = 1, \dots, K, \end{aligned} \tag{3.20}$$

where λ is a Lagrange multiplier, and chosen such that $\sum_k p_k \le \rho$. Note that, with ZF precoding all the users will have the same channel gain and there is no scope of power allocation.

3.4.5 Complexity Analysis

In the sub-section 3.3.4, we have discussed complexity of the QR-base beam selection algorithm is $\sum_{i=0}^{N-K-1} (N-i)\mathcal{O}(2(N-i)K^2)$. Similarly, complexity of the MWM-based beam selection algorithm is $\mathcal{O}(NK^2)$ [111, 112], whereas greedy beam selection algorithm does not perform any mathematical operation. It is just a iteration based search algorithm. Thus, greedy beam selection is an algorithm with a very low computational complexity, i.e, $\mathcal{O}(K)$. It is worth noting that the complexity of greedy and MWM-based beam selection algorithms are providing satisfactory performance with low complexity than "QR-based" beam selection algorithm.

3.4.6 Numerical Results

We present the performance of the proposed algorithms through Monte Carlo simulations, by considering two different mmWave beamspace MU-MIMO downlink

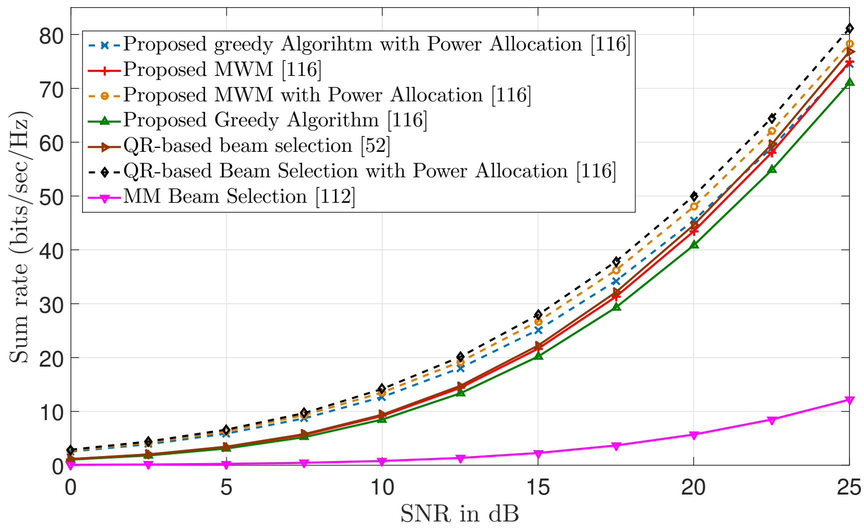

Figure 3.4 Sum rate performance of different beam selection algorithms for a sparse system, with $N = 256$, and $K = 32$

systems: A sparse system, with $N = 256$ beams and $K = 32$ users, and a dense system with $N = 64$ beams and $K = 32$ users. Each user has a single receive antenna. We consider one LoS component with $\alpha_k^{(0)} \sim \mathcal{CN}(0,1)$ and two NLoS components with $\alpha_k^{(1)} \sim \mathcal{CN}(0,10^{-2})$ and $\alpha_k^{(2)} \sim \mathcal{CN}(0,10^{-3})$, from the antenna array at AP to k^{th} user. Further, $\alpha_k^{(\ell)}$ are mutually uncorrelated. The spatial angles, $\theta_k^{(\ell)}$, $\forall k, \forall \ell$, are chosen randomly with a uniform distribution over the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$ and independent of each other.

Fig. 3.4 depicts the sum rate achieved by beam selection algorithms in the sparse system. "QR-based" beam selection refers to the algorithm proposed in [53] and shown to outperform all the existing algorithms including the "IA" beam selection [110] and the full dimensional ZF [113]. It is observed that the performance of MWM beam selection is very close to that of the "QR-based" beam selection. The "QR-based" beam selection is an iterative algorithm that computes QR decomposition of an $(N-i) \times K$ matrix in i^{th} iteration, $i = 0, \ldots, (N-K-1)$, resulting in an overall complexity of $\sum_{i=0}^{N-K-1}(N-i)\mathcal{O}(2(N-i)K^2)$. With a complexity of only $\mathcal{O}(NK^2)$, the "MWM" beam selection becomes an attractive alternative: without considerable loss in the sum rate performance, "MWM" saves significantly in computational complexity. The greedy algorithm, which is a much simpler algorithm compared to "MWM", results in a comparable sum rate performance. Thus,

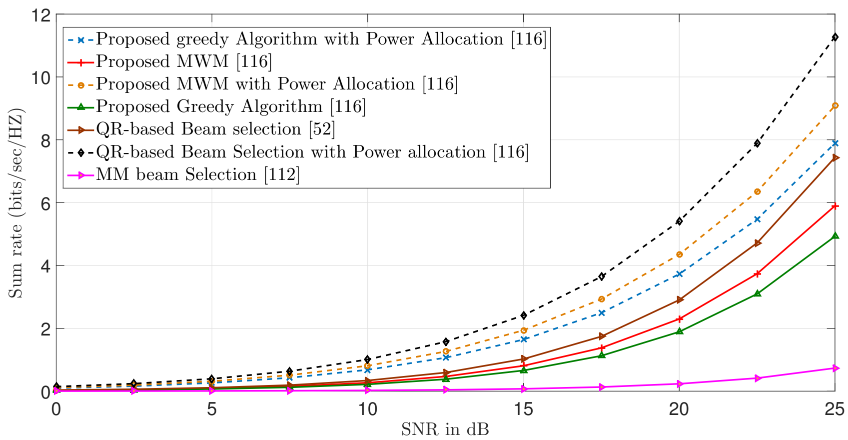

Figure 3.5 Sum rate performance of different beam selection algorithms for a sparse system, with $N = 64$, and $K = 32$

if we have stringent limitations on the affordable computational complexity, greedy algorithm provides us with a better choice for beam selection.

Fig. 3.5 shows that in a dense system, the performance of the existing and the proposed beam selection algorithms are compared. Similarly, it is observed as when the system was sparse (i.e., Fig. 3.4). The performance gap between the "QR-based" beam selection and the proposed algorithms is more noticeable. Thus, in a dense system, to enjoy the benefits of low complexity of the proposed algorithms, one has to sacrifice noticeable loss in the sum rate.

3.5 Concluding Remarks

In this chapter, mmWave beamspace MU-MIMO downlink system is considered. To reduce the required number of beams, a "QR-based" beam selection algorithm is discussed. Further, low complexity algorithms, i.e., a greedy and "MWM-based" beam selection is discussed while taking advantages of mmWave cellular system with regards to the availability of users in the cell and cell size. Next, the performance metrics (spectral efficiency and energy efficiency) are discussed, and compared with the other existing beam selection algorithms and also complexity of the proposed beam selection algorithms are also discussed, and compared with the existing beam selection algorithms. It has been observed that the "QR-based" beam selection algorithm outperforms the existing beam selection algorithms. On the other hand, a greedy and "MWM-based" beam selection algorithms achieve performance comparable to the "QR-based" beam selection algorithm with very low complexity.