Low-Complexity Algorithm for Full-Array Antenna Selection based on Matching Pursuit

Jonathan N. Gois, Tadeu N. Ferreira

Abstract—Massive MIMO systems have received a growing interest since the apogee of 5G and the perspective of 6G. Massive MIMO allows an efficient usage of spectral resources and provides higher data rates. When the number of antennas increases, the energy consumption and the hardware cost in the base station also increase. The high energy consumption and processing may be addressed through the selection of antennas. Among various possible structures in the literature, the present work proposes the combination of full-array architecture to make a virtual vectorization and antenna selection using matching pursuit. The algorithm was tested in different channels with complexity reduction in our tests while maintaining bit-error rate in most scenarios.

Keywords—Massive MIMO, Antenna Selection, Matching Pursuit.

I. INTRODUCTION

Massive Multiple-Input Multiple-Output (MIMO) systems have achieved importance by improving the spatial multiplexing gain and system energy efficiency [1]. Some emerging wireless systems have used the millimeter-wave range to achieve higher data throughput in a less occupied frequency range. Nevertheless, millimeter wave frequencies are affected by high path loss, scattering, and penetration loss [2].

To overcome a higher path loss in the millimeter-wave frequency range, a large array of antennas can be used to increase the directionality of the beam. In spite of these benefits, the large number of antennas also results in a large number of radio-frequencies (RF) chains, and a high cost to estimate all channels between each antenna and terminal [3].

In a massive MIMO system, using all available antennas provides a higher spectral efficiency. On the other hand, this approach results in a maximum energy consumption. While still taking advantage of this system, a possible manner to save energy is to use only some of the antennas for transmitting.

In the literature, some architectures are used to transmit with a large number of antennas. Fully connected architectures are characterized by the connection of each antenna element to the RF chain [1]. This approach provides a good spectral efficiency in spite of the high complexity. On the other hand, we have partially connected architectures that are defined by the connection of groups of antennas into the same RF chain. In this scheme, some antennas are connected to simplify the choice of the transmitting set [4].

One of the methods that exploits the group connection to a single RF chain is the antenna selection method proposed by Gharavi-Alkhansari et al. [5], which decomposes the channel capacity expression in the contribution of each antenna. Optimal and Sub-optimal solutions suggested by Gao et al. [6] also decompose the channel capacity performing additional steps to improve their results.

For both aforementioned structures [5], [6], increasing the number of antennas also results in an increase in the number of RF chains, which could be costly. To reduce the required hardware of the RF chain, some architectures of massive MIMO systems share the RF chain with a group of antennas to reduce the cost. Then, to reduce the complexity and the costs, the virtual vectorization [3] is proposed to adjust the full-array antennas into a reduced-complexity architecture without any hardware improvement.

The traditional methods of antenna selection (AS) focus on optimizing the channel capacity constrained to a limited power of each antenna. This can be treated as a convex optimization problem [7], but its solution is considered computationally expensive. To reduce the computational cost, a family of greedy methods which exploits the sparsity structure of the transmitting antenna vector has been used [8], [9], [10]. The use of these methods varies according to the architecture. Those methods decompose a representation of the entire antenna set in sparse components [11].

With the aim of reducing energy consumption, Ubiali et al. [12] propose an energy-efficient flexible antenna selection method. The Fixed Sub-array Selection (FSS) [12] divides the antenna array into disjoint sub-arrays and maximizes the channel capacity according to the chosen subset. An idea used in this work is based on the adjustable-flexible antenna selection (FAS). By varying the number of selected antennas along the time, it is possible to reduce the energy consumption.

In this paper, we propose a low-complexity AS algorithm based on a Matching Pursuit (MP) channel decomposition [9]. The joint implementation of virtual vectorization and greedy algorithm reduces the computational cost. To test the proposed Sub-array Matching Pursuit Antenna Selection (SM-PAS) method, it is evaluated in different channels and obtain a reduction of computational complexity in comparison to benchmark algorithms with a close bit-error rate.

Section II describe the system modeling used in the deduction of the algorithms, the neighboring cells do not generate
significant co-channel interference (CCI) in the interest cell. The downlink system is composed of \( M \) antennas and \( K \) terminals. In full-array architecture, each antenna is linked with an RF chain. The power \( \eta \) used in each antenna may be controlled. In some of the simulated scenarios, the target BS has knowledge of the downlink Channel State Information (CSI). In others, there is only an estimate of the channel between each antenna for each receiver. With this estimate, a precoder can be defined. The precoder is responsible to mitigate the effects of the channel in the symbol stream between neighbors cells and the close users.

In the transmission model, we are interested in sending only one symbol for each terminal. The symbol message can be written as \( q \in \mathbb{C}^{K \times 1} \). To minimize the channel and user effects in this symbol, we consider the precoding before the transmission. In this work, it was performed using a linear precoding scheme as [9]:

\[
    x = P \text{diag}(\eta) (1/2) q, \tag{1}
\]

where \( P \in \mathbb{C}^{M \times K} \) is the linear precoding matrix and \( \text{diag}(\eta) \) is the power allocation for each terminal. Here, we use the same transmission power for all terminals. The normalized energy [13] for each terminal is given by \( 1/K \). Using this formulation, the received signal in each terminal can be written as:

\[
    y_k = g_k x + \omega_k. \tag{2}
\]

In each terminal, the received message \( y_k \) is built by the channel response between all \( M \) antennas and the \( k^{th} \) terminal \( (g_k \in \mathbb{C}^{M \times 1}) \) applied to the sent signal. Besides the channel effect, each terminal is struck by additive noise \( \omega_k \). Each \( \omega_k \) is a realization of circularly symmetric complex Gaussian distribution \( \mathcal{CN}(0, \sigma^2) \).

Equation (2) can be rewritten in the matrix form for all terminals. Firstly, we will define the full channel matrix between all \( M \) antennas and the \( K \) terminal as follows:

\[
    G_{M \times K} = \begin{bmatrix}
        g_{11} & g_{12} & \cdots & g_{1K} \\
        \vdots & \ddots & \vdots & \vdots \\
        g_{M1} & g_{M2} & \cdots & g_{MK}
    \end{bmatrix}, \tag{3}
\]

\[
    = \begin{bmatrix}
        g_1 & g_2 & \cdots & g_K
    \end{bmatrix}. \tag{4}
\]

Then, the received signal for all \( K \) terminals is given by:

\[
    y = G^T x + w, \tag{5}
\]

where \( w \) is the noise vector. We can also define the received signal according to the symbol message \( q \) and the precoding matrix \( P \):

\[
    y = G^T P q + w. \tag{6}
\]

In this work, we use the zero-forcing (ZF) [14] scheme. The ZF precoding matrix is defined as:

\[
    P_{M \times K} = G^* (G^T G^*)^{-1} \tag{7}
\]

In order to illustrate the described system, Figure 1 shows the full-array architecture. Instead of a vector message \( q \), we use a block data stream \( K \times N_S \). The received block data can be computed as the aforementioned description.

---

A. System with Antenna Selection

Using all available antennas, the maximum energy consumption is achieved. To exploit the redundancy of spatial diversity of massive MIMO antenna array, we can activate only \( S \) out of \( M \) antennas to save power. The respective channel matrix response is described as \( G_s \in \mathbb{C}^{S \times K} \).

The choice of activated antennas can be made using several criteria, including [6]:

1) Exhaustive search over all antennas:
   Performing all combinations of activated antennas, we can find the best subset that results according to the desired metric. However, this solution is infeasible.

2) Maximizing the MIMO Channel Capacity [15]
   The MIMO channel capacity is given by:
   \[
   C(G_s) = \log_{2} \det(I_S + G_s^H \eta G_s). \tag{8}
   \]

   By optimizing Equation (8), we can achieve the best throughput with \( S \) selected antennas.

3) Sparse Recovery:
   In this approach, the main idea is to construct a sparse dictionary and select the best antennas according to the input dictionary.

   We denote the selected or not antennas using the vector \( z^{1 \times M} (z_i \in \{0, 1\}) \). If the \( i^{th} \)-antenna was selected, then \( z_i \) equals 1. Eliminating the disconnected antennas from full-array, we can describe the reduced channel matrix \( G_s \) as:

\[
    G_{s}^{S \times K} = \text{rem}(\text{diag}(z^{1 \times M}) G) \tag{9}
\]

\[
    = \begin{bmatrix}
        g_{11} & g_{12} & \cdots & g_{1K} \\
        \vdots & \ddots & \vdots & \vdots \\
        g_{S1} & g_{M2} & \cdots & g_{MK}
    \end{bmatrix}. \tag{10}
\]
where $S \leq K$.

With a smaller number of antennas, the precoded data $x$ is also reduced. The power for each antenna is scaled by the number of terminals, so the reduced data $x$ also represents the energy saving. Therefore, we can write the new received signal as:

$$ y = G_s^T P_s q + w $$

(11)

where $P_s$ is the zero-forcing precoding matrix considering the reduced MIMO channel $G_s$.

III. GREEDY ANTENNA SELECTION

In order to evaluate the antenna selection algorithm, the literature shows some methods to compute without an exhaustive search. Instead of selecting $S$ out of $M$ antennas and compute the channel capacity (Equation 8) for all $\binom{M}{S}$ possibilities, we can remove the contribution of each antenna from the capacity of the channel and choose the most relevant ones [6].

In other words, this procedure consists in decomposing a representation of the entire channel model into known components. If we choose a representative dictionary for our problem, the matching pursuit algorithm can find the best projection of $b$ in there. The vector $b$ must also be defined as representative for the entire array. Considering $z$ the array that indicates which antennas are selected, the problem to be solved is [9], [10]:

$$ \min_{z \in \mathbb{C}^{M \times 1}} \|Dz - b\|_2^2, $$

(12)

subject to $\|z\|_0 = S$.

The dictionary $D$ must be sparse and overcomplete [16] in order to represent the best set of elements in $z$ that minimizes the $\ell_2$-norm as described in Equation 12. In this approach, the dictionary has the function to adapt the vector $b$ in the problem context. Therefore, an appropriate choice for the dictionary is essential. Gharavi-Alkhansari et al. [5] suggested the pair $D$ and $b$ as:

$$ D = \left( I_M - G^H G \right)^{-1} $$

(13)

$$ b_j = g_j^H g_j $$

(14)

Rather than only one vector $b$, the authors proposed a set of vectors in order to be decomposed in the dictionary. The dictionary is part of the channel capacity expression given by Equation (8).

Mendonça et al. [9] propose an antenna selection algorithm denoted as ZF-GAS (Zero-Forcing Greedy Antenna Selection) that minimizes the distance between only the vector that represents the full-array MIMO channel $b$ and the dictionary. In their work, the dictionary $D$ was defined as:

$$ D_{ZF} = \left[ \text{vec}(p_1^P), \text{vec}(p_2^P), \ldots, \text{vec}(p_M^P) \right] $$

(15)

where each $p_i$ is a column of the zero-forcing precoding matrix $P$. The operator vec{ } is responsible to resize the matrix $P \in \mathbb{C}^{M \times K}$ in to a $MK \times 1$ vector. For the same reasons that the dictionary was defined in this form, the vector $b$ is defined as:

$$ b = \text{vec}(P^T P) $$

(16)

Using this representation, only one decomposition should be performed, reducing considerably the number of operations to select the antennas in the BS.

IV. CHANNEL-LEVEL PROPOSED METHOD

The proposed method was based on the Reduced-Complexity Architecture [4]. When we reduce the number of RF Chains, the computational cost of the chosen configuration decreases, and the energy is less used in the BS. However, the spectral efficiency also decreases. The maximum spectral efficiency is achieved with all antennas selected.

The arrangement in the Figure 2 is our target to select subsets of antennas. In order to obtain a flexible architecture, Molisch et al. [3] show a structure named Virtual Sectorization. The joint choice of antennas reduces the computational cost of the choice, that is, it saves energy and gives flexibility to the entire structure.

The Greedy Antenna Selection proves to be computationally efficient [9] and gives similar results when compared to others in terms of error metrics. To perform a flexible algorithm with the same characteristics as ZF-GAS, we propose to divide the entire array into subsets. Firstly, the search occurs in the same way as in ZF-GAS. The principal decomposition of the vector $b$ is found in the dictionary $D$. Both variables are constructed equally described in the Equations 15 and 16. To virtually sectorize the array, when the $i$th antenna is selected, the adjacent antennas are also selected, according to definition of the sub-array size ($S_i$). The contribution of that dictionary position is removed from vector $b$ as described in the Algorithm 1.

Algorithm 1 Sub-array Matching Pursuit Antenna Selection - SMPAS

- Input Parameters: $G$, $D_{ZF}$, $b$, $S_i$ and $S_s$.
- Loop into Sub-array Space:
  - for $(i \leftarrow 1 ; i \leq M ; i \leftarrow i + S_i)$ do
  - max_ip $\leftarrow -\infty$
  - closest_id $\leftarrow 0$ $\triangleright$ Index of the best Antenna Selection
  - Matching Pursuit Algorithm:
    - $j \leftarrow 1$
    - for $(col \leftarrow I(1); I$ is not empty; $col \leftarrow$ next index ) do
      - ip $\leftarrow (r(j), D_{ZF}(col))$
      - if ip $> max_ip$ then
        - max_ip $\leftarrow$ ip
        - closest_id $\leftarrow$ col
      - end if
  - end for
- Update Residue:
  - $\kappa \leftarrow closest_id - \text{mod}(closest_id, S_s)$
  - $r(j + 1) \leftarrow r(j) - \sum_{k=0}^{\kappa} D_{ZF}(k)$
  - $j \leftarrow j + 1$
- Update Valid Indexes:
  - Remove used indexes from $I$.
- end for

V. RESULTS

A. General Parameters

Concerning the evaluation tests, the environment is defined with $M = 128$ antennas disposed in a full-array architecture serving $K = 16$ terminals/users. Sent data is encoded with BPSK (Binary Pulse Shift-Keying) and the binary data is produced randomly. Each block of message has length equals to $N_s = 200$. The data is transmitted in a channel imposed by Gaussian noise, as referenced in Section II varying the SNR.

3
(signal-to-noise ratio) between -15 dB and 5 dB or 10 dB and 40 dB, according the channel. A Monte Carlo simulation was performed to estimate the bit error rate (BER) average per user. The number of Monte Carlo runs was fixed at 500, resulting in a total of $10^5$ bits tested for each user. With respect to the proposed algorithm, we defined the sub-array size $S_s = \{1,2,4,8,16\}$.

**B. Simulation Scenario**

As in any communication system, the previous system model is completely dependent on the channel. Here, we analyze the algorithm using two different channels. The first one is an uncorrelated Rayleigh fading channel, widely used in the literature. Basically, the channel is a realization of a circular normal random variable independent and identically distributed (i.i.d.) for each antenna. To perform a realistic scenario, the SMPAS was also tested in the DeepMIMO Dataset [17]. The used scenario in this work is referred by the dataset as “O1_60”, which is an outdoor scenario. It is composed by two main streets with one intersection. The main street is 600 meters long and 40 meters wide, the other one is 440 meters long and 40 meters wide. In each street, they positioned 18 base stations and more than 1 million possible users. The buildings along the street have different shapes and height. To complete the simulation, various materials were used in the ray-tracing to determine the response for each antenna and each user. In our test, the users are located around 10 meters from the BS. A summary of the used parameters in this simulation were described in the Table I.

**TABLE I: DeepMIMO channel parameters used in the simulations.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Scenario Name</td>
<td>O1_60</td>
</tr>
<tr>
<td>Central Frequency</td>
<td>60 GHz</td>
</tr>
<tr>
<td>Active Base Station</td>
<td>BS-3</td>
</tr>
<tr>
<td>Antenna Spacing</td>
<td>0.5λ</td>
</tr>
<tr>
<td>Geometry of the Array</td>
<td>ULA (Uniform Linear Array)</td>
</tr>
<tr>
<td>Number of OFDM Subcarriers</td>
<td>1</td>
</tr>
<tr>
<td>Number of paths</td>
<td>5</td>
</tr>
<tr>
<td>Number of Antennas</td>
<td>$M = 128$</td>
</tr>
<tr>
<td>Number of Terminals</td>
<td>$K = 16$</td>
</tr>
</tbody>
</table>

**C. Perfect Channel Knowledge**

Considering a CSI perfect estimation, we applied the proposed method in both aforementioned channels.

In Figure 3, the Rayleigh channel was employed. In a simple environment, the SMPAS achieves almost the same bit-error performance, for all sub-array sizes. An important note is the decreasing number of operations on the ascending size of the sub-array in a direct proportion. To illustrate the full-complexity algorithm, the ZF-GAS is performed when the sub-array size ($S_s$) equals 1.

The SMPAS is also tested in the DeepMIMO Channel previously described. In Figure 4, for $S_s = 2$, that is, with 50% fewer operations and 50% less energy used, a slight loss of performance is observed.

**D. Imperfect Channel Estimation**

Using an imperfect CSI knowledge, we test our algorithm by applying an MMSE channel estimation. We consider an uplink channel with 15 dB of SNR for Rayleigh and 30 dB of SNR for DeepMIMO and its estimation is used as the estimated downlink CSI. The results are shown in the Figure 5 and 6.

The difference in the SNRs imposed on uplink channels is motivated by the estimation impairment due to the more severe characteristics of time and frequency dispersion of a Deep MIMO channel. With a rough estimation, we can observe that Rayleigh channel maintains the BER of a configuration without sub-arrays.

Nevertheless, for the DeepMIMO channel, the error imposed by the channel estimation completely impairs the antenna selection algorithm. When we increase the sub-array size, the bit-error rate increases even further.

We realize that the proposed algorithm presents a very good compromise between bit-error rate and computational complexity for the tested range of sub-array size for exact CSI. In the estimated CSI scenario, for a Rayleigh channel, the BER remains the same. For a DeepMIMO channel, there is a good compromise for $S_s = 2$, whereas the bit-error rate is excessive for other sizes of sub-arrays.

**VI. CONCLUSION**

In this paper, we propose a low-complexity algorithm based...
on matching pursuit for full-array antenna selection to apply in massive MIMO systems. The concept of an array composed by sub-arrays was virtually implemented using virtual sectorization. An extension of ZF-GAS [9] was suggested to reduce the number of operations in the base station. Moreover, with the reduction of necessary power to transmit, the base station also saves energy.

The present paper brings two types of channel models. The Rayleigh channel and a realistic scenario using DeepMIMO tool [17] to perform the ray-tracing between each antenna and the user. DeepMIMO allows imposing more difficulties and a different scenario for this type of algorithm.

Since the proposed algorithm is non-optimum, the measured BER is also underperforming when compared to other methods. However, using the lowest sub-array value, it is possible to reduce computational cost by 50% with also reducing energy consumption. With an imperfect CSI, the performance of the system with sub-arrays presents a similar bit-error rate for a Rayleigh channel, whereas there is a higher degradation for DeepMIMO channels.

At last, the present paper demonstrates the possibility to extend the virtual sectorization in existing full-arrayarchite-
tures to save energy reducing the operations performed in the BS. An extension of this work for other greedy algorithms is also considered as a future work, as well as, using a non-linear precoding to address DeepMIMO channels.

**ACKNOWLEDGEMENT**

The authors thank Marcele O. K. de Mendonça, M.Sc., for providing some of the scripts used in this work and FAPERJ Projects E-26/010.001579/2019, E-26/210.524/2019 (248292) for funding.

**REFERENCES**


