XL-MIMO Energy-Efficient Antenna Selection under Non-Stationary Channels

Gabriel Avanzi Ubiali, Taufik Abrão

Abstract

Massive multiple-input-multiple-output (M-MIMO) is a key technology for 5G networks. Within this research area, new types of deployment are arising, such as the extremely-large regime (XL-MIMO), where the antenna array at the base station (BS) has extreme dimensions. As a consequence, spatial non-stationary properties appear as the users see only a portion of the antenna array, which is called visibility region (VR). In this challenging transmission-reception scenario, an algorithm to select the appropriate antenna-elements for processing the received signal of a given user in the uplink (UL), as well as to transmit the signal of this user during downlink (DL) is proposed. The advantage of not using all the available antenna-elements at the BS is the computational burden and circuit power consumption reduction, improving the energy efficiency (EE) substantially. Numerical results demonstrate that one can increase the EE without compromising considerably the spectral efficiency (SE). Under few active users scenario, the performance of the XL-MIMO system shows that the EE is maximized using less than 20% of the antenna-elements of the array, without compromising the SE severely.

Index Terms

Extremely-large antenna regime; Non-Stationary Channels; Antenna selection; Energy efficiency; spectral efficiency; precoding; combining

I. INTRODUCTION

Massive multiple-input-multiple-output (M-MIMO) is one of the key technologies for 5G networks [1], which permits that more than one single user transmits simultaneously with high spectral and energy efficiencies and using the same spectrum, i.e., many antennas simultaneously serve many users using the same time-frequency resource [2]. In MIMO networks, the base station (BS) estimates the channel coefficients and employs a transmit precoding scheme in the downlink (DL) and a receive...
combining scheme in the uplink (UL), giving each user a different spatial signature [3], [4]. M-MIMO wireless communication is a special case of MIMO systems using hundreds of antennas at the BS, providing sufficient spatial dimensions to uncover the fundamental properties of M-MIMO: channel hardening, large array gain and asymptotic inter-terminal channel orthogonality (favorable propagation) [5]. Thus, it can provide large improvements over traditional systems in both energy and spectral efficiencies.

As the number of BS antennas increases, it is possible to focus the transmission and reception of signal energy into ever-smaller regions of space, which brings huge improvements in throughput and EE. However, it may come with the computational complexity increasing, as well as with the increasing of implementation cost and power consumption, what advocates for new deployments that take real advantage from increasing the number of BS antennas to the order of hundreds or thousands, without severe problems due to the holdbacks above cited. Moreover, in order to make real advantage from the deployment of such a large number of antenna elements, it is desirable to distribute them over a substantially large area in order to increase the antenna separation and coverage [6]. One potential approach is the extremely-large MIMO (XL-MIMO) regime, where the antenna array is integrated into large building structures [5].

When a moderate number of (several tens of) antennas is compactly deployed in the BS, the entire array will receive approximately the same amount of energy from each user, i.e., the channel is spatially stationary [7]. On the other hand, in the XL-MIMO regime, different parts of the array may observe the same propagation paths with varying power and phases or distinct propagation paths. Then, the majority of energy received from a specific user concentrates on small portions of the entire array, which is a channel property called spatial non-stationarity, which has been observed by recent channel measurements [8]–[10] and can be introduced in the channel model by using the concept of visibility region (VR) [5], [11]–[13]. However, it is worth saying that the density of VRs influences on the size of the portion of the array that the user can see. For instance, if the VR density is sufficiently high, all portions of the array are able to receive some signal energy.

The conventional M-MIMO signal processing architecture is centralized at the BS, what means that the signals are received at the BS (UL) and transmitted by the BS (DL) deploying all elements of antennas. Then, the associated computational complexity becomes a challenge when employing extremely large arrays, specially in crowded scenarios, due to the need to transfer excessively large amounts of data received by the array to the processing unit [7]. A promising solution is to use only a portion of the whole element-antennas array to perform receive combining as well as the transmit precoding to each user.

Hence, by appropriately selecting a subset of BS antennas to communicate with each user, the receiver is able to capture almost the totality of the energy transmitted by that user, while reducing the
interference coming from the other users and therefore potentially increasing the spectral efficiency (SE). Furthermore, by having higher SE and reduced power consumption, one can obtain higher EE. It is important to highlight that there is a growing concern about how to improve the EE in wireless communications, as the increasing data rates and the increasing number of users connected to the network increase substantially the overall energy consumption [14]. For this reason, in this work we take both SE and EE as performance metrics to be analyzed in section V. Finally, one can summarize the benefits of the VR-based subarray antenna selection architecture as: a) computational complexity reduction; b) overall energy saving by activating a reduced number of antenna-elements; c) as a result an EE increasing; d) while potentially improving the overall system SE by selecting appropriate antenna-elements associated to the each user VR.

Considering that the antenna array experiences spatial non-stationarities when using large-aperture arrays, in [15], authors propose two new channel estimation methods that, besides estimating the channel vector, obtains the position of the scatterers and the visibility regions, which may be useful for transceiver design. In [11], authors study the impact of spatial non-stationarity where the channel energy is concentrated on a portion of the array, i.e., the VR, in terms of signal-to-interference-plus-noise ratio (SINR) performance. In [5], authors show that, when M-MIMO systems operate in extra-large scale regime, several important MIMO design aspects change, due to spatial non-stationarities, where the users see only a portion of the array and, inside the VR, different parts of the array see different propagation paths. Moreover, three low-complexity data detection algorithms are proposed in [12] as candidates for uplink communication in XL-MIMO systems.

In [7], authors designed an efficient detector for extra-large-scale massive MIMO systems with the subarray-based processing architecture, by extending the application of the expectation propagation principle. Their analysis is based on bit error rate (BER) performance. A different approach to the antenna selection methodology proposed herein, in [16], authors propose a design based on machine learning to select a small portion of the array that contains the beamforming energy to the user, aiming to overcome the prohibitive complexity of XL-MIMO systems. They provide numerical results in terms of sum-rate performance. To the best of author’s knowledge, this is the first work addressing the subarray-based processing architecture, which remarkably provides huge reduction in the computational complexity, through the point of view of improving the overall system EE.

**Contribution.** We deal with an XL-MIMO system equipped with a subarray-based processing architecture, in order to reduce the overall system computational complexity. In such scenario, we propose a novel algorithm to judiciously select the antenna-elements subarray that will communicate with each user, aiming at obtaining higher EE while reducing the power consumption, when compared to the whole antenna array activation to communicate with every user. Thus, the contribution of this work can be summarized as: i) we propose an antenna selection procedure to improve simultaneously the
overall system EE by reducing the power consumption, taking into account the spatial non-stationarity assumption while taking advantage of the VRs features; ii) the proposed algorithm also provides a considerable computational complexity reduction; iii) a comprehensive analysis development on how the proposed procedure impacts the system performance is developed, highlighting and characterizing its benefits when comparing to the condition of using the entire antenna array to communicate with every user.

The remainder of the paper is organized as follows. The adopted XL-MIMO channel model and the channel estimation procedure is developed in Section II; this section also provides the ergodic UL and DL spectral efficiencies expressions, based on the signal-to-interference-plus-noise ratio (SINR). Section III focuses on the proposed antenna selection procedure, as well as on the computational complexity aspects. The EE definition and a detailed circuit power model are discussed in Section IV. Section V examines numerical results corroborating our findings, while the main conclusions are presented in Section VI.

II. SYSTEM MODEL

We consider the UL and the DL of a single-cell multiuser XL-MIMO system with an $M$-antenna BS and $K$ single-antenna users at each cell, operating over a bandwidth of $B$ Hz. The channel estimates are acquired via UL synchronous pilot transmission. The time-division duplex (TDD) operation mode was chosen because of its advantages over the frequency-division duplex (FDD) mode. TDD does not require quantized channel state information (CSI) to be sent by the BS to the user via feedback, because of channel reciprocity, avoiding excessive overhead [3], [17].

The channel coherence time ($T_C$) is divided into UL pilot, UL data and DL data transmission, as Figure 1 shows. The number of symbols that fits in a channel coherence block is $\tau_c = T_CB_C$, being $B_C$ the coherence bandwidth [2]. In order to estimate the channel, each of the $K$ users of a given cell is assigned a different pilot sequence. There are $\tau_c$ symbols per coherence block, of which $\tau_p$ are dedicated to UL pilot transmission, $\tau_u = \epsilon_u(\tau_c - \tau_p)$ symbols are dedicated to UL data transmission and $\tau_d = \epsilon_d(\tau_c - \tau_p)$ symbols are dedicated to DL data transmission, where $\epsilon_u + \epsilon_d = 1$. The number of available orthogonal pilot sequences is equal to its length ($\tau_p$). As we need $K$ sequences, we can take $\tau_p = K$. Thus, the time required for pilots is proportional to the number of users served. The number of users that can be served is therefore limited by the coherence time, which itself depends on the mobility of the users [2].

The antenna-elements are uniformly spaced over a uniform linear $L$-length array containing $M$ elements ($M$-ULA). The coordinates of the first and the $M$-th antennas are $(0,0)$ and $(L,0)$, respectively, which means that, when $M \geq 1$, the spacing between the antennas is $\frac{L}{M-1}$. The users are placed over a rectangle that extends along the antenna array in one dimension and between
Figure 1: Each coherence block is divided into the UL pilot transmission for channel estimation purpose, the UL data transmission, and the DL data transmission.

A minimum \((d_{\text{min}})\) and a maximum distance \((d_{\text{max}})\) in the other dimension, following a uniform distribution over this area. The coordinates of the \(m\)-th antenna are denoted by \((a_m, 0)\), and its distance to the \(k\)-th user is denoted by \(d_{mk}\). A typical system configuration is represented by Figure 2.

Figure 2: Typical system spatial configuration.

A. XL-MIMO Channel Model

In XL-MIMO systems, spatial non-stationarities occur frequently since typically \(d_{\text{max}} < L\), what means that a given user probably sees only a small part of the antenna array. The propagation environment contains numerous objects reflecting the signal, which are called scattering points. Each scattering point has an associated VR. Herein, we assume that each user can see the antenna array through \(N_c\) different VRs. The \(i\)-th VR that is visible to the \(k\)-th user extends from the \(c_{ik}\)-th to the \((c_{ik} + N_{ik})\)-th antenna and is denoted by \(C_{ik} = \{c_{ik}, c_{ik} + 1, \ldots, c_{ik} + N_{ik}\}\). These VRs may overlap, and the set \(C_k = C_{1,k} \cup \ldots \cup C_{N_c,k}\) contains the indices of the antennas that are visible.
to the $k$-th user. Considering that each user sees the antenna array through more than one VR, it is a way of taking into account that not only one subset of the array containing contiguous antenna elements may be visible. Indeed, the mobile user may see more than one portion of the array. For instance, if $M = 100$ in Figure 2, it may see antennas 10 to 20 and 25 to 40, which means there is something in the propagation environment preventing the signal from reaching antennas 21 to 24.

In general, the distances between a given user and all the BS antennas are considered to be the same. However, in the XL-MIMO scenario, as the length of the antenna array is not negligible, the pathloss varies throughout the array, mainly when $d_{\text{max}} \ll L$. The pathloss coefficient between the $m$-th antenna and the $k$-th user is given by:

$$b_{mk} = \frac{b_0}{(d_{mk})^\gamma}$$

(1)

where $\gamma \geq 2$ is the pathloss exponent and $b_0$ determines the median channel gain at a reference distance of 1 m [6]. The parameters $b_0$ and $\gamma$ are functions of the carrier frequency, antenna gains, and vertical heights of the antennas, which are derived from fitting (1) to measurements [6]. Finally, the channel vector $h_k = [h_{1k} \ldots h_{ Mk}]^T$ between the BS and the $k$-th user is given by:

$$h_k = a_k \odot \sqrt{b_k} \odot h_k$$

(2)

where $b_k$ is a vector whose $m$-th element is $b_{mk}$, and $h_k$ is the independent Rayleigh fading component, which accounts for the short-scale fading and follows a complex-Gaussian distribution $h_k \sim \mathcal{CN}(0_M, I_M)$. The coefficient $a_{mk} = [a_k]_m$ indicates whether the $m$-th antenna is visible to the $k$-th user ($a_{mk} = 1$) or not ($a_{mk} = 0$), and is given by:

$$a_{mk} = \begin{cases} 
1 & m \in \mathcal{C}_k \\
0 & \text{otherwise}. 
\end{cases}$$

(3)

B. UL Pilot Transmission

We assume the channel is estimated via UL synchronous pilot transmission, which means that all users simultaneously send pilot sequences from the same pilot codebook. They have length $\tau_p$ and form an orthogonal set. Herein, it is assumed that each user is assigned a different pilot sequence. Then, the pilot sequences’ set is $\Psi = [\psi_1 \ldots \psi_K] \in \mathbb{C}^{\tau_p \times K}$ and the orthogonality condition states that $\Psi^H \Psi = \tau_p I_{\tau_p}$, i.e.:

$$\psi_i^H \psi_k = \begin{cases} 
\tau_p & i = k, \\
0 & i \neq k.
\end{cases}$$

(4)

During the UL pilot transmission, the $k$-th user transmits the pilot sequence $\psi_k \in \mathbb{C}^{\tau_p}$, with transmit power $p_p$. The elements of $\psi_k$ are scaled by $\sqrt{p_p}$, forming the signal $s_k = \sqrt{p_p} \psi_k^H$, to be
transmitted over $\tau_p$ UL samples. As a result, the BS receives the signal $Y^p \in \mathbb{C}^{M \times \tau_p}$:

$$ Y^p = \sum_{i=1}^{K} \sqrt{\tau_p} h_i \psi_i^H + \mathbf{N}^p $$

(5)

where $\mathbf{N}^p \in \mathbb{C}^{M \times \tau_p}$ is the noise matrix at the receiver of the BS with i.i.d. entries following a complex normal distribution with zero mean and variance $\sigma_{UL}^2$.

As the information about which antennas are visible for each user is unknown, it might be necessary to consider obtaining the channel estimates by using estimators that require no prior statistical information, such as the least-squares (LS). Moreover, as we consider that each user is assigned a different pilot sequence and these sequences are mutually orthogonal, there is no pilot contamination. Hence, the imperfections on the channel estimates are just due to the noise power at the BS antennas during the UL pilot transmission. It advocates for using LS channel estimation rather than MMSE.

The LS estimate of $h_k$ is attained by [6]:

$$ \hat{h}_k = \frac{1}{\tau_p \sqrt{P_p}} Y^p \psi_k $$

(6a)

$$ = h_k + \frac{1}{\tau_p \sqrt{P_p}} \mathbf{N}^p \psi_k $$

(6b)

The last term in (6b) is the equivalent noise vector, which adds imperfections to the channel estimates and follows a complex normal distribution: $\mathbf{N}^p \psi_k \sim \mathcal{CN}(0_M, \tau_p \sigma_{UL}^2 I_M)$. Finally, the estimated channel matrix is $\hat{\mathbf{H}} = [\hat{h}_1 \ldots \hat{h}_K]$, while $\mathbf{H} = [h_1 \ldots h_K] \in \mathbb{C}^{M \times K}$ is the true channel matrix. According to equation (6a), the channel estimation process corresponds to the inner product of $M$ complex vectors of length $\tau_p$, requiring $M \tau_p$ multiplications between complex numbers (or $3M \tau_p$ multiplications between real numbers) to estimate the channel vector of each of the $K$ users. Herein, we consider that both multiplication and division between real numbers correspond to 1 floating-point operation (flop). As the channel estimation process is performed once per coherence block, its computational complexity, defined in number of flops per coherence block [fpcb], is:

$$ C_{CE} = 3MK\tau_p $$

(7)

**C. UL Data Transmission**

The received signal $\mathbf{r} \in \mathbb{C}^{M}$ at the BS during the UL data transmission is:

$$ \mathbf{r} = \sum_{k=1}^{K} h_k x_k + \mathbf{n} $$

(8)

---

1 Consider $x = a + jb$ and $y = c + jd$. The hardware implementation of the complex multiplication $xy = ac - bd + j[(a + b)(c + d) - ac - bd]$ involves 3 real multiplications and 5 real sums. Only those will be considered, due to their very greater hardware complexity compared to the real sum operation.
where $x_k$ is the signal sent by the $k$-th user and $n \sim \mathcal{CN}(0, \sigma^2_{UL} I_M)$ contains the noise received at the BS antennas. By utilizing a suitable combining vector, $v_k \in \mathbb{C}^M$, the BS detects the $k$-th user’s signal as follows:

$$y_k = v_k^H r$$  \hspace{1cm} (9)

Herein we consider the two simplest types of linear processing for receive combining: zero-forcing (ZF) and maximum-ratio (MR), which are respectively defined by:

$$V = \hat{H}(\hat{H}^H \hat{H})^{-1}$$  \hspace{1cm} (10)

and

$$V = \hat{H}$$  \hspace{1cm} (11)

where the matrix $V = [v_1 \ldots v_K]$ is the collection of the combining vectors. ZF induces considerably smaller intra-cell interference than MR, yielding significantly better performance under interference-limited conditions, which is normally the case. On the other hand, it increases computational complexity significantly when employing large antenna arrays, due to multiplications of complex numbers, and when serving a great number of users, due to the size of the $K \times K$ matrix that is inverted [3], [6]. Due to these particularities, it is interesting to compare the system performance in terms of SE and EE considering such linear processing techniques.

The channels are practically constant within a coherence block, while the signals and noise take new realization at every sample. Then, the instantaneous SINR is actually an expectation over one coherence block, what means that $p_{UL}^k = \mathbb{E}(|x_k|^2)$ - which is the UL transmit power of the $k$th user - and $\sigma^2_{UL}$ will be taken instead of the instantaneous values of $|x_k|^2$ and $|n_k|^2$, respectively. Thus, one can define the SINR of the $k$th user during the UL data transmission as

$$\gamma_{UL}^k = \frac{p_{UL}^k |v_k^H h_k|^2}{\sum_{i=1}^{K} p_{UL}^i |v_k^H h_i|^2 + \sigma^2_{UL} ||v_k||^2}$$  \hspace{1cm} (12)

As a result, the UL ergodic spectral efficiency is defined by [6]:

$$\text{SE}_{UL} = \frac{\tau_u}{\tau_c} \sum_{k=1}^{K} \mathbb{E}\{\log_2(1 + \gamma_{UL}^k)\}$$  \hspace{1cm} (13)

**D. DL Data Transmission**

In the DL, the information to be transmitted by the BS to the $k$-th user, $x_k$, needs to be precoded, by using the precoding vector $w_k$. The signal to be transmitted, denoted by $s \in \mathbb{C}^M$, is generated
as:

\[ s = \sum_{k=1}^{K} w_k x_k \]  

(14)

We denote the matrix containing the collection of the precoding vectors by \( \mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_K] \in \mathbb{C}^{M \times K} \). The UL-DL duality motivates a simple precoding design principle: selecting the DL precoding vectors as the normalized version of their respective combining vectors [6]:

\[ \mathbf{w}_k = \frac{\mathbf{v}_k^*}{||\mathbf{v}_k||} \]  

(15)

Assuming no receive combining, user \( k \) receives the signal:

\[ y_k = \mathbf{h}_k^T s + n_k \]  

(16)

where the received noise follows the distribution \( n_k \sim \mathcal{CN}(0, \sigma_{DL}^2) \). Analogously to (12) for the UL, the SINR of the \( k \)th user in the DL data transmission is defined as:

\[ \gamma_{DL}^k = \frac{p_{DL}^k |\mathbf{h}_k^T \mathbf{w}_k|^2}{\sum_{i=1}^{K} p_{DL}^i |\mathbf{h}_k^T \mathbf{w}_i|^2 + \sigma_{DL}^2} \]  

(17)

where \( p_{DL}^k \) is the downlink transmit power assigned for user \( k \), i.e., \( \sqrt{p_{DL}^k} \) scales the vector \( \mathbf{w}_k \), which has unit norm. Finally, the DL ergodic spectral efficiency is given by [6]:

\[ \text{SE}_{DL} = \frac{\tau_d}{\tau_c} \sum_{k=1}^{K} \mathbb{E}\{ \log_2(1 + \gamma_{DL}^k) \} \]  

(18)

### III. Antenna Selection for Combining and Precoding in XL-MIMO

In this section, we propose an algorithm to select the antenna-elements in an XL-MIMO system for received signal processing (combiner) of a given user during the UL and transmit the signal (precoder) of this user during the DL. One advantage of not using all the \( M \) available antennas is the reduction of the computational complexity and the circuit power consumption (Tx and Rx operations), as less antennas are active at the same time. Furthermore, the throughput potentially increases, because the interference power at the receivers decreases since each antenna individually does not serve all the \( K \) users simultaneously.

#### A. HRNP-based Antenna Selection Criterion and Algorithm

First, Algorithm [1] computes the vector \( \theta_k \in \mathbb{C}^M \), which is a quantitative indicator of the quality
of the channel between the $k$-th user and each of the $M$ antennas as:

$$\theta_{mk} = \frac{|\hat{h}_{mk}|^2}{\sum_{i \neq k}^K |\hat{h}_{mi}|^2}, \quad m = 1, 2, \ldots, M$$

(19)

where $\hat{h}_{mk} = [\hat{h}_k]_m$ and $\theta_{mk} = [\theta_k]_m$. A high signal intensity may be obtained when $|\hat{h}_{mk}|^2$ is strong.

Selecting the $N$ strongest $\theta_{mk}$ values among $m = 1, \ldots, M$ in (19) for each user, provides the highest received normalized power (HRNP) antenna selection criterion. On the other hand, the terms $|\hat{h}_{mi}|^2$, $i \neq k$, are related to the interference intensity. Higher $\theta_{mk}$ values are therefore associated to higher SINRs on the signal detection, as defined by the equations (12) and (17), and consequently higher SE and EE.

Second, the Algorithm 1 obtains the set $D_k$ (lines 5–10), which contains the indices of the $N$ antennas with the highest $\theta_{mk}$ values. Only these $N$ antenna-elements are activated for user $k$.

Lastly, Algorithm 1 computes the receive combining and the transmit precoding vectors of the $k$-th user (lines 11–14) based on the matrix $\hat{H}_k \in \mathbb{C}^{N \times K}$, which contains all the columns of the estimated channel $\hat{H}$ but only the rows corresponding to the elements of the set $D_k$.

As the set $D_k$ contains the indices of the antennas that are active for the $k$th user, the superset $\mathcal{D} = \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_K$ contains all the indices of the antennas that are active for any user. The number of elements in $\mathcal{D}$, denoted by $N_{\text{act}}$, corresponds to the total number of active elements of antenna.

Notice that the rows of the combining and the precoding matrices corresponding to the antennas whose indices are not in the set $D_k$ are set equal zero.

---

**Algorithm 1** Antenna selection (AS) for receive combining and transmit precoding

*Input:* $M$, $N$, $K$, $\hat{H}$  
*Output:* $V$, $W$

1. Initialize the combining matrix $V$ with $0_{M \times K}$
2. for $k = 1$ to $K$ do
3. Compute vector $\theta_k$ via eq. (19)
4. Reinitialize the set of the indices of the antennas: $\mathcal{M} = \{1, \ldots, M\}$
5. Initialize $D_k = \emptyset$
6. for $n = 1$ to $N$ do
7. find $m^* = \arg \max_{m \in \mathcal{M}} \theta_{mk}$
8. $\mathcal{M} = \mathcal{M} \setminus m^*$
9. $D_k = D_k \cup \{m^*\}$
10. end for
11. $\hat{H}_k = \hat{H}(D_k, :)$
12. If MR is selected: $V_{MR}(D_k, k) = \hat{H}_k(:, k)$
13. If ZF is selected: $V_{ZF}(D_k, k) = [\hat{H}_k(\hat{H}_k^H\hat{H}_k)^{-1}](:, k)$
14. $W(:, k) = \frac{V(:, k)^*}{||V(:, k)||}$
15. end for
B. Computational Complexity

We first address the complexity of computing the ZF combining vectors (line 13, Algorithm 1). Recalling that $\mathbf{H}_k \in \mathbb{C}^{N \times K}$, the multiplication of $\mathbf{H}_k^H$ by $\mathbf{H}_k$ requires $\frac{K^3+KN}{2}$ complex multiplications\footnote{Being $\mathbf{A} \in \mathbb{C}^{a \times b}$ and $\mathbf{B} \in \mathbb{C}^{b \times c}$, the multiplication $\mathbf{AB}$ requires $ac$ inner products between $b$-length vector, what corresponds to $abc$ complex multiplications. However, if $\mathbf{B} = \mathbf{A}^H$, the Hermitian symmetry is utilized. Thus, only the $a$ diagonal elements of $\mathbf{A} \cdot \mathbf{B}$ and half of the $a^2 - a$ off-diagonal elements need to be computed, what gives $\frac{a^2+ab}{2}$ complex multiplications \cite{6}.}. using the Hermitian symmetry. When the inverse of a matrix is multiplied by another matrix, the $\text{LDL}^H$ decomposition can be used to achieve an efficient hardware implementation \cite{6}. The decomposition of $\mathbf{H}_k^H \mathbf{H}_k$ requires $\frac{K^2-K}{3}$ complex multiplications \cite{6}. Finally, we need to multiply the matrix $\mathbf{H}_k$ by the $k$-th column of the matrix $(\mathbf{H}_k^H \mathbf{H}_k)^{-1}$, which requires $KN$ complex multiplications plus $K$ complex divisions to compute $\mathbf{D}^{-1}$ \cite{6}, \cite{13}. Considering complex multiplications and complex divisions to correspond to 3 and 7 flops\footnote{Considering $x = a + jb$ and $y = c + jd$, then $\frac{x}{y} = \frac{ac+bd}{y^2}$, while the computation of $xy^*$ requires 3 real multiplications. The computation of $|y|^2 = c^2 + d^2$ requires 2 real multiplications. Finally, the complex division $\frac{x}{|y|^2}$ corresponds to 2 real divisions, making a total of 7 real operations.}, respectively, the computation of the combining vector $\mathbf{v}_k$ has a complexity of $3 \left( \frac{K^2+K}{2}N + \frac{K^2-K}{3} + KN \right) + 7K$ flops per coherence block. Thus, the computational complexity to obtain the whole combining matrix $\mathbf{V}$ is given by:

$$
C^{\text{UL-ZF}}_{\text{SP-C}} = K^4 + \frac{3}{2}K^3N + \frac{9}{2}K^2N + 6K^2 \quad \text{[fpcb]} \tag{20}
$$

As defined in \cite{11}, MR combining does not require multiplications or divisions, because it is given directly from the channel estimates (from Algorithm \cite{11} one can see that the $k$-th column of the MR combining matrix is simply a copy of the $k$-th column of $\mathbf{H}_k$). However, in practical implementations, we typically normalize the combining vector such that $\mathbf{v}_k^H \mathbf{h}_k$ in front of the desired signal $x_k$ is close to one. Thus, this normalization requires 1 complex division per user \cite{6}, resulting in a total of $7K$ flops per coherence block. Finally, the complexity of computing the MR combining matrix is given by:

$$
C^{\text{UL-MR}}_{\text{SP-C}} = 7K \quad \text{[fpcb]} \tag{21}
$$

The precoding vectors ($\mathbf{w}_k$) are chosen as the normalized versions of the combining vectors ($\mathbf{v}_k$), as described in equation \cite{15}. The computation of $||\mathbf{v}_k||$ requires $2N$ real multiplications\footnote{Consider the complex vector $\mathbf{x} = [x_1, \ldots, x_N]$. The computation of $||\mathbf{x}|| = \sqrt{\sum_{n=1}^{N} |x_n|^2}$ depends on previously obtaining $|x_m|^2$. Being $x_n = a_n + jb_n$ a complex scalar, $|x_n|^2 = a_n^2 + b_n^2$ requires 2 real multiplications and 1 real sum. Therefore, the computation of $||\mathbf{x}||$ requires $2N$ real multiplications.} and the division of $\mathbf{v}_k$ by $||\mathbf{v}_k||$ also requires $2N$ real divisions\footnote{The division of a complex scalar $x = a + jb$ by a real scalar $c$ requires 2 real divisions. Therefore, the division of $x$ by $||\mathbf{x}||$ requires $2N$ real divisions.}, resulting in a total of $4N$ flops. Thus,
computation of the precoding matrix has a complexity of:

$$C_{SP-C}^{DL} = 4KN \quad \text{[fpcb]} \quad (22)$$

Notice that the rows of the combining and the precoding vectors of the user $k$ that correspond to the antennas whose indices are not in the set $D_k$ are set equal zero. It reduces the complexity to obtain $y_k$, as in (22), because the BS will use $N$ elements of the vectors $v_k$ and $r$, instead of $M$ elements, resulting in $3N$ flops. This procedure is repeated $\tau_u$ times per coherence block. Similarly, the complexity to precode the information during DL, $w_k x_k$, following (14), is reduced since the BS only uses $N$ elements of the vector $w_k$, also resulting in $3N$ flops. This task is performed $\tau_d$ times per coherence block. Finally, the computational complexity associated to the reception and transmission of the information, in number of flops per coherence block, is:

$$C_{SP-R/T} = 3(\tau_u + \tau_d)KN \quad \text{[fpcb]} \quad (23)$$

To obtain $\theta_{mk}$ in (19), the BS computes $2K$ real multiplications and 1 real division. As there are $M$ antennas and $K$ users, the associated complexity is:

$$C_{SP,\theta} = (2K + 1)MK \quad \text{[fpcb]} \quad (24)$$

Finally, the total signal processing computational complexity, in flops per coherence block, when employing MR and ZF processing in the context of XL-MIMO antenna selection is given respectively by:

$$C_{TSP} = C_{SP-C}^{UL-MR} + C_{SP-C}^{DL} + C_{SP-R/T} + C_{SP,\theta} \quad \text{[fpcb]} \quad (25)$$

and

$$C_{TSP} = C_{SP-C}^{UL-ZF} + C_{SP-C}^{DL} + C_{SP-R/T} + C_{SP,\theta} \quad \text{[fpcb]} \quad (26)$$

Hence, if antenna selection (AS) procedure is not applied, the complexities given in (20), (22) and (23) will be higher, since all the $M$ antennas always will be active for all the $K$ users, i.e., $N = M$.

IV. SE AND EE IN XL-MIMO SYSTEMS

The SE is defined as the sum-rate in bits per channel use [bpcu] achieved in the UL + DL, expressed as:

$$SE = SE_{UL} + SE_{DL} \quad \text{[bpcu]} \quad (27)$$
The overall network EE can be defined as the number of bits that can be reliably transmitted per unit of energy, which is the same as the throughput per unit of power \( \frac{\text{bit/s}}{\text{W}} \), given by:
\[
EE = \frac{B \cdot SE}{P_{UL\,TX} + P_{DL\,TX} + P_{tr\,TX} + P_{CP}} \quad \text{[bit/J]}
\]  
(28)
where the denominator includes all power consumption terms required to make the wireless communication system operational. Hence, the term
\[
P_{tr\,TX} = \frac{\tau_p}{\tau_c \eta_{UL}} K p_p
\]  
(29)
accounts for the total power consumed by the power amplifiers during the UL pilot transmission, while
\[
P_{UL\,TX} = \frac{\tau_u}{\tau_c \eta_{UL}} \sum_{k=1}^{K} p_{ULk}
\]  
(30)
and
\[
P_{DL\,TX} = \frac{\tau_d}{\tau_c \eta_{DL}} \sum_{k=1}^{K} p_{DLk}
\]  
(31)
refers to the UL and DL power consumed for data transmission, respectively, being \( \eta_{UL} \) and \( \eta_{DL} \) the power amplifier efficiency at the BS and at the users, respectively. \( P_{CP} \) represents the circuit power consumption. A detailed model for \( P_{CP} \) is discussed in the sequel.

A. Circuit Power Model

The following circuit power consumption model based on [6] is adopted in this work:
\[
P_{CP} = P_{FIX} + P_{TC} + P_{CE} + P_{C/D} + P_{BH} + P_{SP}
\]  
(32)
where \( P_{FIX} \) is a constant quantity. It accounts for the power consumption required for site-cooling, control signaling and load-independent power of backhaul infrastructure and baseband processors. The power consumed by the backhaul is commonly modeled as the sum of two parts: load-independent and load-dependent. The last one will be included in \( P_{BH} \), and is typically the least significant part (around 20%) [6].

The other terms of the model represented in (32) account for the power consumption of the transceiver chains (\( P_{TC} \)), the channel estimation process (\( P_{CE} \)), the channel coding and decoding units (\( P_{C/D} \)), the load-dependent backhaul (\( P_{BH} \)) and the linear processing at the BS (\( P_{LP} \)). Each of these terms depends on at least one of the main system parameters: \( M \), \( K \) and the ergodic spectral efficiency (SE).

The power consumption of the transceiver chains (\( P_{TC} \)) involves the power consumed by the BS local oscillator (\( P_{LO} \)), the power required by the circuit components (converters, mixers and filters)
of each BS antenna ($P_{BS}$) and the power necessary to run the circuit components (mixers, filters, amplifiers and oscillator) of each single-antenna user ($P_{UE}$), as described by the equation

$$P_{TC} = P_{LO} + N_{act}P_{BS} + KP_{UE}$$

(33)

The computational complexity associated to the channel estimation process is given by the equation (7). Hence, the resulting power consumption is given by:

$$P_{CE} = \frac{BC_{CE}}{\tau_{c} L_{BS}}$$

(34)

where $L_{BS}$ is the computational efficiency at the BS, in $[\text{flop/s W}]$. Similarly, the total signal power consumption is given by:

$$P_{SP} = \frac{BC_{TSP}}{\tau_{c} L_{BS}}$$

(35)

The power consumed by the channel coding and decoding units is defined as

$$P_{C/D} = B\, \text{SE} (P_{\text{COD}} + P_{\text{DEC}})$$

(36)

which increases linearly with the actual rates. $P_{\text{COD}}$ and $P_{\text{DEC}}$ are the coding and decoding power densities, respectively, in $[\text{watt/}\text{bit/s}]$. For simplicity, $P_{\text{COD}}$ and $P_{\text{DEC}}$ are assumed to be the same in both UL and DL.

The load-dependent backhaul power consumption, necessary for the UL and DL data transmission between the BS and the core network, is modeled as

$$P_{BH} = B\, \text{SE} \, P_{BT}$$

(37)

where $P_{BT}$ is the backhaul traffic power density, in $[\text{watt/}\text{bit/s}]$. There is also a load-independent backhaul power consumption, which can be included in $P_{\text{FIX}}$.

V. Numerical Results

In the sequel we present numerical results based on Monte-Carlo simulations with the objective to demonstrate that the proposed algorithm provides an EE increase while reducing considerably the computational complexity and the power consumption in the context of XL-MIMO systems. Table\textsuperscript{1} contains a list of the main deployed parameter values, similar to those adopted in [6], [19], [20].

As each coherence block contains $\tau_{c}$ symbols per second, $B/\tau_{c}$ is the number of coherence blocks per second. If a given signal processing has a computational complexity denoted by $C$, representing the number of flops per coherence block, and $L$ is the computational efficiency, representing the number of flops per Joule of energy, then $C/L$ represents the energy consumption per coherence block. Therefore, the associated power consumption is $\frac{BC}{\tau_{c} L_{w}}$. 

\begin{table}
\end{table}
A. Simulation Setup and System Configuration

In our simulations, the antenna array contains $M = 100$ antenna elements and each user sees the array through 3 different VRs, i.e., $N_c = 3$. Furthermore, $N_{ik}$ is taken as a uniform random variable distribution over the interval $[0.1M, 0.3M]$, while the index of the first antenna that is inside this VR, denoted by $c_{ik}$, follows a uniform distribution in the interval $[1, M - N_{ik}]$.

The set $C_k = C_{1,k} \cup \ldots \cup C_{N_c,k}$ contains the indices of the antennas that sees the $k$-th user. Thus, the number of elements in $C_k$ corresponds to the total number of antennas seen by user $k$. We obtained this value numerically from our simulations. As well as the other numerical results presented in the paper, this value was averaged out of 1,000 random realizations, obtaining the average number of active antennas for the $k$-user as $|C| = 55.8$. It means that approximately 56 antennas are seen by each user, on average. Also from our simulations, we found that, when $K = 4$, each antenna sees an average of 2.23 users, while for $K = 40$ this number goes to 22.3, resulting in higher interference levels.

Moreover, we assume equal power allocation (EPA), what means the transmit power is the same for all users. Hence, during the UL data transmission, $p_{UL}^k = p_{UL}$, while during the DL data transmission, the total available transmit power at the BS, $P_{DL}$, is equally divided among the $K$ users, so that the power allocated to the $k$-th user is $p_{DL}^k = P_{DL}/K$. We have adopted $p_p = p_{UL} = 0.1$ W and $P_{DL} = 1$ W. Although the position of the users (and consequently the pathloss), the short-scale fading and the VRs are random variables, the numerical results are statistically relevant because they represent an average over 1,000 realizations.

Figure 3 shows an example of the XL-MIMO system user spatial distribution, where $K = 4$ and $M = 16$, for simplicity. Each triangle represents one of the 16 BS antennas, while each colored circle represents one of the 4 mobile users, which are randomly distributed over a rectangular area. Taking a random channel realization, the portion of the array that each user sees is indicated by the horizontal line with its correspondent color. User 1, for example, sees the antennas 1 to 10, while user 2 sees the antennas 9 to 15, excepting the antenna 13. This fragmentation of the VR into two parts may occur if any object is blocking the signal in that region.

In this example, the algorithm was set up to define $N = 4$ antennas to communicate with each user. In the figure, the triangle that represents a given antenna $m$ is painted with the user $k$ color if user $k$ sees antenna $m$. For example, antennas 1, 2, 3 and 9, which are in blue, were designated to communicate with user 1. Notice that all these 4 antennas are part of the VR of user 1. Although the signal from user 1 probably achieves the antennas 4 to 8 with higher intensity than antenna 9, choosing one of these antennas would increase the received interference power, mainly due to user 3. Observe that antenna 4 is not active, while antenna 14 serves users 2 and 4, simultaneously.
| Parameter                                                   | Value                                      |
|-------------------------------------------------------------|--------------------------------------------|
| Pathloss attenuation exponent: $\gamma$                    | 2.5                                        |
| Median channel gain at a distance of 1 m: $b_0$             | $2.95 \cdot 10^{-4}$                      |
| Number of NLoS VRs for each user: $N_C$                    | 3                                          |
| Antenna array length: $L$                                  | 60 m                                       |
| Number of BS antennas (ULA), $M$                           | 100                                        |
| Number of mobile users, $K$                                | $\{4, 40\}$                               |
| Minimum distance ($d_{\text{min}}$)                       | 5 m                                        |
| Maximum distance ($d_{\text{max}}$)                       | 30 m                                       |
| Number of Monte Carlo realizations: $T$                    | 1,000                                      |
| Transmission bandwidth: $B$                                | 20 MHz                                     |
| Channel coherence bandwidth: $B_C$                         | 100 kHz                                    |
| Channel coherence time: $T_C$                              | 2 ms                                       |
| Total UL noise power: $\sigma_{UL}^2$                     | $-100$ dBm                                 |
| Total DL noise power: $\sigma_{DL}^2$                     | $-80$ dBm                                  |
| UL pilot transmit power per user: $p_p$                     | 0.1 W                                      |
| UL data transmit power per user: $p_{UL}$                  | 0.1 W                                      |
| Total DL transmit power: $P_{DL}$                           | 1.0 W                                      |
| Fraction of UL transmission: $\epsilon_u$                  | 0.4                                        |
| Fraction of DL transmission: $\epsilon_d$                 | 0.6                                        |
| Power amplifier efficiency at the users: $\eta_u^{UL}$     | 0.5                                        |
| Power amplifier efficiency at the BSs: $\eta_u^{DL}$       | 0.4                                        |
| Computational efficiency at the BS: $L_{BS}$                | 75 Gflop/s W                               |
| Fixed power consumption: $P_{\text{FIX}}$                  | 10 W                                       |
| Power consumed by local oscillators at BS: $P_{\text{SYN}}$| 0.2 W                                      |
| Power consumed by circuit components at BS: $P_{\text{BS}}$| 0.2 W                                      |
| Power consumed by circuit components at UE: $P_{\text{UE}}$| 0.2 W                                      |
| Power density for coding of data signals: $P_{\text{COD}}$| $0.1$ W Gbit/$^2$/s                        |
| Power density for decoding of data signals: $P_{\text{DEC}}$| $0.8$ W Gbit/$^2$/s                      |
| Power density for backhaul traffic: $P_{\text{BT}}$        | $0.25$ W Gbit/$^2$/s                      |

Figure 3: Example of a XL-MIMO system, with 4 users being served by a BS equipped with a 16-antenna linear antenna array. The figure illustrates each user’s VR and the antennas that the proposed algorithm designates to communicate with each user.
Therefore, considering that only one antenna was designated to communicate with 2 users at the same time, we can say that the algorithm succeeded in avoiding the interference from other users to affect the SINR while reducing the computational complexity.

B. Dependency of the System Performance on $N$

In the following, we present numerical results demonstrating how the number of active antennas per user ($N$) influences the system performance, in terms of throughput and EE, and the resource consumption, in terms of computational complexity and power consumption. From these results, we can see the advantages of appropriately selecting the antennas subset (AS strategy) against utilizing the whole antenna array strategy to serve all users at the same time (no-AS strategy). To refine the comprehension upon these results, which depend on the processing scheme (MR or ZF), Fig. 4 provides, for two different scenarios ($K = 4$ and $K = 40$), valuable insights on how $N$ influences the received signal, interference and noise power, during the UL and during the DL, which are given by the definitions in Table II.

Table II: Definitions considered in Fig. 4 in [dBm].

| Definition | Equation |
|------------|----------|
| **Definition 1**: UL average received signal power | $S_{UL} = 10 \cdot \log_{10} \left( \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\{p_{UL}^{k} |\mathbf{v}_{k}^{H}\mathbf{h}_{k}|^2\} \right) + 30$ |
| **Definition 2**: UL average received interference power | $I_{UL} = 10 \cdot \log_{10} \left( \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1 \atop i \neq k}^{K} \mathbb{E}\{p_{UL}^{i} |\mathbf{v}_{k}^{H}\mathbf{h}_{i}|^2\} \right) + 30$ |
| **Definition 3**: UL average received noise power | $N_{UL} = 10 \cdot \log_{10} \left( \frac{\sigma_{UL}^2}{K} \sum_{k=1}^{K} \mathbb{E}\{|\mathbf{v}_{k}|^2\} \right) + 30$ |
| **Definition 4**: DL average received signal power | $S_{DL} = 10 \cdot \log_{10} \left( \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\{p_{DL}^{k} |\mathbf{h}_{k}^{H}\mathbf{w}_{k}|^2\} \right) + 30$ |
| **Definition 5**: DL average received interference power | $I_{DL} = 10 \cdot \log_{10} \left( \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1 \atop i \neq k}^{K} \mathbb{E}\{p_{DL}^{i} |\mathbf{h}_{k}^{H}\mathbf{w}_{i}|^2\} \right) + 30$ |
| **Definition 6**: DL average received noise power | $N_{DL} = 10 \cdot \log_{10} (\sigma_{DL}^2) + 30$ |

Figures 4a and 4b compare the UL signal, interference and noise power, as given by the definitions.
Def. 1 to 3, when the number of users is 4 and 40, respectively. When $N < K$, it is not possible to execute ZF, due to singularity problems with the matrix $\hat{H}^H \hat{H}$. First, one can see that the received power levels of ZF are higher than of MR. The reason is that ZF receive combining is the pseudoinverse of $\hat{H}$ (recall that the elements of the channel matrix include the pathloss effect), while MR combining is simply a copy of $\hat{H}$. Second, by comparing Figures 4a and 4b, one can observe that, as expected, when $K = 40$, the received interference power ($I_{UL}$) is higher than when $K = 4$, because the selectivity of the receive combining deteriorates. Third, looking at the MR curves, one can verify that, by selecting more antennas to be active for each user, the received signal power increases, but so does the interference and noise power. It does not occur when employing ZF combining. By increasing $N$, ZF performs better at reducing the interference and noise power, because the receive combining becomes more selective. Fourth, if the channel estimates are reliable, ZF combining will force the average received signal power ($S_{UL}$) to $p_{UL}$, as $|\gamma_k^H h_k|^2 = 1$, from the definition of ZF combining. It can be observed in the figure, what attests that the channel estimation predicted by eq. (6a) provides good estimates.

Unlike the UL case, during the DL (see Fig. 4c and 4d), the average received noise, $N_{DL}$, does not depend on the precoding scheme (compare Def. 3 and 6). Another important point is that the desired signal power ($S_{DL}$) is the same magnitude order for both MR and ZF, because the precoding vectors are normalized, unlike the combining vectors. Finally, as the total available DL transmit power ($P_{DL}$) is distributed among the $K$ users, the individual DL transmit power ($p_{DL}$) is inversely proportional to $K$. That is why the average signal and interference power are smaller in Fig. 4d than in Fig. 4c. The exception is $I_{DL}$ when employing MR precoding, which is less efficient than ZF at eliminating the interference.

Fig. 4e and 4f depict the UL, DL and UL+DL throughput, given by $B \cdot SE_{UL}$, $B \cdot SE_{DL}$ and $B \cdot SE$, respectively. The throughput resulting from the no-AS strategy can be obtained in the point $N = 100$. Thus, we can see that in a scenario with few active users, the AS strategy improves considerably the system throughput when using MR. On the other side, considering the $K = 40$ scenario, the high interference levels deteriorate the selectivity of MR combining and precoding, and the AS strategy cannot improve the throughput. The same behavior is observed with ZF processing, independently of the number of users. However, as the array is physically large and the users have a VR corresponding to around 50% of the array, on average (as discussed in the beginning of Section V-A), increasing $N$ beyond $M/2$ can barely improves the throughput. Thus, by taking $N$ close to $M/2$, we can obtain almost the same throughput achieved with the no-AS strategy, as shown in Figures 4e and 4f, while benefiting from lower computational complexity and consequently lower power consumption.
Figure 4: Average received power of the desired signal, undesired signal (interference) and noise as given by Def. 1 to 6), as well as the system throughput, during the UL and the DL data transmission. Two scenarios are considered: $K = 4$ and $K = 40$ users.
C. Computational Complexity

While $D_k$ contains the indices of the $N$ antennas that are active for user $k$, the set $D$ contains the indices of antennas $N_{\text{act}}$ that are active for any of the $K$ users. Fig. 5 addresses the dependency of $N_{\text{act}}$ on $N$. In the scenario with 4 users, the whole antenna array is expected to be active when $N \geq 34$, approximately. When $K = 40$, $N_{\text{act}}$ scales faster than when $K = 4$, and reaches the limit of 100 when $N \approx 8$. If a small $N$ is sufficient to substantially increase the EE or throughput, one will benefit from the reduced circuit power consumption, as $P_{\text{TC}}$ is proportional to $N_{\text{act}}$, according to (33). It can be evidenced by Fig. 6c and 6d that with small values of $N$, the AS strategy provides considerable reduction on the power consumption comparing to the entire antenna array activation (no-AS strategy).

![Figure 5: Average number of active antennas per user ($N$) versus the total number of active BS antennas ($N_{\text{act}}$).](image)

Table III provides a quantitative analysis of the impact of the proposed HRNP-based AS scheme on the overall XL-MIMO computational complexity, given by $C = C_{\text{CE}} + C_{\text{TSP}}$. The complexity associated to the no-AS strategy can be obtained from (20), (21), (22) and (23) by simply replacing $N$ with $M$ in these equations, and cancelling $C_{\text{SP,θ}}$ out in (25) and (26). The complexities are in the unit of [flop/s], which is given by $BC/\tau_c$, recalling that $C$ is measured in flop per coherence block [fpcb]. The first three columns of the table define four scenarios, with different values of $M$, $N$ and $K$. The other columns contain the complexity associated to each processing scheme (MR and ZF), when using or not the AS algorithm. Notice that the algorithm yields very lower complexities, which is particularly advantageous in high system dimensions (many BS antennas and many users). Also, although providing lower throughput than ZF, one can benefit from using MR due to its much
smaller complexity. However, in the fourth scenario, the AS’s complexity surpasses the no-AS’s when employing MR, because \( C_{SP,\theta} \) is the only term of \( C_{TSP} \) that is proportional to \( K^2 \). It does not occur with ZF, as the term \( C_{UL-ZF-SP,\theta} \) is much more significative than \( C_{SP,\theta} \).

Table III: AS computational complexity in [Gflop/s], discriminated by processing scheme.

| \( M \) | \( N \) | \( K \) | no-AS MR | AS MR | no-AS ZF | AS ZF |
|-------|-------|-------|----------|-------|----------|-------|
| 32    | 4     | 2     | 4        | 1     | 4        | 1     |
| 128   | 16    | 8     | 62       | 12    | 76       | 14    |
| 512   | 64    | 32    | 990      | 369   | 3,848    | 819   |
| 2048  | 256   | 128   | 15,834   | 17,551| 702,031  | 126,822|

Fig. 6a shows the linear dependency of the computational complexity on \( N \) and the remarkable computational complexity reduction provided by the adopted AS strategy. As a consequence, there is also a reduction in the total power consumption, defined as \( P_{tot} = P_{UL,TX} + P_{DL,TX} + P_{TX} + P_{CP} \), as depicted in Fig. 6c. Moreover, Fig. 6b reveals a significant complexity increase when the number of active users grows from 4 to 40. Furthermore, by avoiding all antennas to be simultaneously active, the adopted HRNP-AS strategy in eq. (19) reduces the transceiver chains power consumption\(^7\), and consequently the total power consumption, as indicated in Fig. 6c and 6d. As a final remark on the advantage in adopting the HRNP-AS strategy is the operation point where the AS power consumption curves meet the no-AS curves is very close to the point where the average number of active antennas meet \( M \) in Fig. 5.

D. Energy Efficiency

Fig. 7 confirms the EE improvement when AS strategy based on highest received power is adopted. Under reduced loading scenario (\( K = 4 \)), the EE is maximized when \( N = 6 \) and \( N = 1 \) for ZF and MR, respectively. It demonstrates that the HRNP-AS strategy can guarantee huge EE improvements in a scenario with few users, while reducing considerably the power consumption and the computational burden, as discussed above. On the other hand, results in Fig. 7b demonstrate that, the AS strategy cannot improve the EE considerably when XL-MIMO operates under high loading scenarios (\( K = 40 \)). However, the HRNP-based AS strategy is still advantageous over the no-AS strategy, as it still provides a considerable complexity reduction. Thus, regardless of the number of users, it is not reasonable to use the whole antenna array to serve all users if about half of the antennas is sufficient to achieve a remarkable EE increasing with a lower computational complexity.

\(^7\)Notice that from (33), power consumption is linearly dependent on \( N_{act} \).
Figure 6: The average consumption of computational and power resources raises when increasing $N$, while depends on the adopted AS strategy.

Figure 7: Energy efficiency as a function of $N$, in both low and high loading scenarios: $K = 4$ and $K = 40$. 
E. Overall XL-MIMO Performance Comparison

Table IV shows the optimal number of selected antennas \(N^*\) that maximizes the EE, when employing both linear MR or ZF filtering. According to the data, \(N^*\) is strongly influenced by the number of active users \(K\). As a rule of thumb, when the number of active users is up to a limit, namely \(K_{\text{max}}\), the EE is maximized when the AS algorithm selects less than 10\% of the antenna array to communicate with each user, \textit{i.e.}, \(N^* \leq M/10\). From the data in Table IV, we see that \(K_{\text{max}} = 20\) for MR and \(K_{\text{max}} = 6\) for ZF.

Table IV: Optimal number of selected antennas for maximizing the EE \((N^*)\) \textit{versus} the number of active users.

| \(K\) | 2 | 4 | 6 | 8 | 10 | 12 | 16 | 20 | 24 | 32 | 40 | 50 |
|------|---|---|---|---|----|----|----|----|----|----|----|----|
| MR   | 3 | 1 | 1 | 1 | 1   | 1   | 1   | 1   | 1   | 48 | 48 | 48 |
| ZF   | 5 | 6 | 8 | 64| 62 | 60 | 58 | 57 | 57 | 60 | 65 | 75 |

Figs. 8a, 8b, 8c and 8d depict the EE, throughput, power consumption and computational complexity, respectively, for the proposed HRNP-AS and no-AS strategy assuming \(N = N^*\). If the system operates under the bound \(K \leq K_{\text{max}}\), the HRNP-AS strategy provides an increasing in the EE and simultaneously reduces the power consumption and the computational burden, while the throughput is very close to the obtained with the no-AS strategy.

Besides, when increasing \(K\) until it is close to \(K_{\text{max}}\), the EE gradually decreases until it meets the no-AS strategy EE curve. It occurs because taking \(N \leq M/10\) is no longer enough to mitigate the interference. Besides, from the point \(K = K_{\text{max}}\), \(N^*\) suddenly jumps to about 50 or 60\% of the ULA array size, \(M = 100\). Finally, when the XL-MIMO system operates over the maximum number of user bound, \(K > K_{\text{max}}\), the AS-strategy is no longer able to improve the EE and simultaneously reduce the power consumption considerably. However, it is still advantageous when compared to the all-antennas activation strategy, as it is still able to reduce the computational burden.

VI. Conclusion

In XL-MIMO systems operating under non-stationary channels, the users see only a portion of the antenna array and the majority of the energy sent by the users is concentrated on this part of the array. Therefore, by appropriately selecting a subset of antennas that communicate to each user, it can be guaranteed to capture almost the totality of the incident energy from that user, while reducing substantially the interference coming from the other \((K - 1)\) users. As corroborated by extensive numerical results, the HRNP antenna selection criterion reduces considerably the complexity of computing the combiners and precoders and consequently reducing the power consumption also.
Furthermore, when the AS-HRNP algorithm is set to select only a few $N$ antennas per user, there may be several antennas which are not activated, reducing the transceiver chains power consumption. As the HRNP-based AS strategy results in almost the same spectral efficiency as the no-AS strategy and a considerable substantial power consumption reduction, as a result, the EE is increased significantly. Furthermore, the extensive numerical results demonstrated the existence of an optimal value of $N$ in terms of maximizing the EE, which depends on the number of users and array size. Also, it is not even advantageous to increase $N$ beyond this optimal value, since neither the throughput nor the EE would be considerably improved while computational burden and energy consumption increases remarkably.

**References**

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5g,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.
[2] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[3] T. Marzetta, E. Larsson, H. Yang, and H. Ngo, *Fundamentals of Massive MIMO*, ser. Fundamentals of Massive MIMO. Cambridge University Press, 2016. [Online]. Available: https://books.google.com.br/books?id=IM0iDQAAQBAJ

[4] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge University Press, 2014.

[5] E. de Carvalho, A. Ali, A. Amiri, M. Angelichinoski, and R. W. H. Jr., “Non-stationarities in extra-large scale massive MIMO,” *CoRR*, vol. abs/1903.03085, 2019. [Online]. Available: http://arxiv.org/abs/1903.03085

[6] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[7] H. Wang, A. Kosasih, C. Wen, S. Jin, and W. Hardjawana, “Expectation propagation detector for extra-large scale massive mimo,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2036–2051, 2020.

[8] S. Hou, Y. Wang, T. Zeng, and S. Wu, “Sparse channel estimation for spatial non-stationary massive mimo channels,” *IEEE Communications Letters*, vol. 24, no. 3, pp. 681–684, 2020.

[9] J. Li, B. Ai, R. He, M. Yang, Z. Zhong, Y. Hao, and G. Shi, “The 3d spatial non-stationarity and spherical wavefront in massive mimo channel measurement,” in *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2018, pp. 1–6.

[10] J. Chen, X. Yin, X. Cai, and S. Wang, “Measurement-based massive mimo channel modeling for outdoor los and nlos environments,” *IEEE Access*, vol. 5, pp. 2126–2140, 2017.

[11] A. Ali, E. D. Carvalho, and R. W. Heath, “Linear receivers in non-stationary massive mimo channels with visibility regions,” *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 885–888, June 2019.

[12] A. Amiri, M. Angelichinoski, E. de Carvalho, and R. W. Heath, “Extremely large aperture massive mimo: Low complexity receiver architectures,” in *2018 IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–6.

[13] J. Flordelis, X. Li, O. Edfors, and F. Tufvesson, “Massive mimo extensions to the cost 2100 channel model: Modeling and validation,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 380–394, 2020.

[14] H. Wang, A. Kosasih, C. Wen, S. Jin, and W. Hardjawana, “Low complexity channel estimation for massive mimo systems,” *IEEE Communications Letters*, vol. 19, no. 1, pp. 120–123, 2015.

[15] J. Chen, X. Yin, X. Cai, and S. Wang, “Measurement-based massive mimo channel modeling for outdoor los and nlos environments,” *IEEE Access*, vol. 5, pp. 2126–2140, 2017.

[16] A. Amiri, C. N. Manch’on, and E. de Carvalho, “Deep learning based spatial user mapping on extra large mimo arrays.” 2020.

[17] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, “Massive mimo performance—tdd versus fdd: What do measurements say?” *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2247–2261, April 2018.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[19] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, “Optimal design of energy-efficient multi-user mimo systems: Is massive mimo the answer?” *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, June 2015.

[20] J. C. M. Filho, C. Panazio, T. Abrão, and S. Tomasin, “Total energy efficiency of tr-mrc and fd-mrc receivers for massive mimo uplink,” *IEEE Systems Journal*, vol. 13, no. 3, pp. 2285–2296, 2019.