Low RF-Complexity Millimeter-Wave Beamspace-MIMO Systems by Beam Selection

Pierluigi V. Amadori, *Student Member, IEEE*, and Christos Masouros, *Senior Member, IEEE*

***Abstract*—Communications in millimeter-wave (mm-wave) spectrum (30–300 GHz) have experienced a continuous increase in relevance for short-range, high-capacity wireless links, because of the wider bandwidths they are able to provide. In this work, we introduce a new mm-wave frequency transmission scheme that exploits a combination of the concepts of beamspace multi-input multi-output (B-MIMO) communications and beam selection to provide near-optimal performances with a low hardware-complexity transceiver. While large-scale MIMO approaches in mm-wave are affected by high dimensional signal space that increases considerably both complexity and costs of the system, the proposed scheme is able to achieve near-optimal performances with a reduced radio-frequency (RF) complexity thanks to beam selection. We evaluate the advantages of the proposed scheme via capacity computations, comparisons of numbers of RF chains required and by studying the trade-off between spectral and power efficiency. Our analytical and simulation results show that the proposed scheme is capable of offering a significant reduction in RF complexity with a realistic low-cost approach, for a given performance. In particular, we show that the proposed beam selection algorithms achieve higher power efficiencies than a full system where all beams are utilized.**

***Index Terms*—High dimensional MIMO, multiuser MIMO, beam selection, capacity maximization, SINR maximization, beamforming.**

I. Introduction

In THE last years, data traffic has suffered an exponential growth, supported by the increasing popularity of mobile devices, such as smartphones and tablets. As a consequence, recent studies predicted that the global mobile data traffic will reach a 66% annual growth rate in the next years [1]. Recent works showed that millimeter-wave (mm-wave) communications offer a promising approach [2]–[5] for meeting this demand. Millimeter-wave technologies, operating in the 30–300 GHz spectrum, are in fact able to provide very high-dimensional MIMO operations and spatial re-use of the spectrum at close distance. The main disadvantage might be identified in the attenuation characteristics [6], [7], because it is widely known that the free space propagation loss is inversely proportional to the wavelength λ. However, a shorter wavelength allows to use more antennas in the same physical space. Therefore, if we consider the same aperture area for two systems working at two different carrier frequencies, the system with a shorter wavelength will allow beamforming with higher gains than the system with a longer wavelength. In fact, it has been demonstrated [8] that a mm-wave mobile broadband system could achieve gigabit per second data rates at distances up to 1 km in an urban environment.

A large or massive MIMO approach in mm-wave frequencies is still prohibitive because of the number of antennas and the high transceiver complexity [9], [10]. In fact, the use of an increased number of antennas leads to an equally large number of radio frequency (RF) chains. At the same time, recent studies [11] proved that systems that involve massive antenna arrays in MIMO are particularly subjected to RF chain imperfections, which are accounted for additional degradations in performance. Moreover it is understood that RF components may consume up to 70% of the total transceiver power consumption [12].

In order to exploit the favorable characteristics of mm-wave frequency communications, research is focusing on the development of new techniques that aim to reduce the hardware complexity of very high dimensional MIMO systems. Previous works tackled the hardware complexity with antenna selection [13]–[15], amongst many others, but showed high degradation in performances compared to the full system. Antenna selection techniques generally require more power from the amplifier in the output stage to compensate the transfer attenuation of RF switches [13] and are characterized by lower values of average SNR [15]. In addition, [14] showed that waterfilling for a full system provides higher performances than the antenna selection at the transmitter and, at the same time, highlighted the costs of selections applied to the transmitter in terms of SNR.

In this paper, we propose an architecture that combines B-MIMO communication [16] with beam selection criteria, based on concepts for antenna selection. We show that the use of beam selection algorithms reduces significantly the RF complexity of the system. The benefits of the system simplification provided by the beam selection are strongly enhanced by the use of Discrete Lens Array (DLA) [17] where there is a direct correspondence between the number of selected beams and radio frequency (RF) chains. Discrete Lens Arrays behave as a convex lens, directing the signals towards different points of the focal surface [18]. Moreover, the fact that with DLA technology narrow beam-widths are preserved even in the reduced RF-chain operation allows to reduce significantly the power required per stream as well as the interference between the streams. As a consequence, the low hardware complexity and the high antenna

Manuscript received September 1, 2014; revised February 26, 2015; accepted April 29, 2015. Date of publication May 8, 2015; date of current version June 12, 2015. This work was supported by the Royal Academy of Engineering, U.K. and the Engineering and Physical Sciences Research Council (EPSRC) project EP/M014150/1. The associate editor coordinating the review of this paper and approving it for publication was G. Abreu.

The authors are with Department of Electronic and Electrical Engineering—University College London, London WC1E 7JE, U.K. (e-mail: pierluigi.amadori@gmail.com; chris.masouros@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2015.2431266

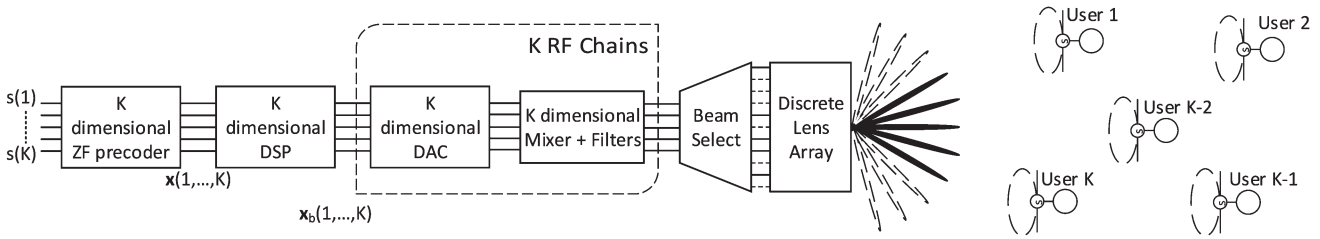

Fig. 1. Proposed scheme block diagram.

gain properties of DLA systems make this technology particularly suitable for mm-wave communications [5]. Therefore, unlike antenna selection, the performance obtained with beam selection in B-MIMO is close-to-optimal, as its application does not affect the beamforming properties at the transmitter. As regards the transmit processing involved, capacity achieving non-linear precoders [19]–[26] are known to involve prohibitive computational complexity. Computationally efficient alternatives are provided by linear precoding [27]–[31], which become near-optimal in massive MIMO systems. Accordingly, in this paper we focus on zero-forcing (ZF) precoders [32], [33], as a practical closed-form precoding solution [34]. In particular, we show how system performances are affected when we use different beam selection algorithms to reduce the RF complexity of the transmitter.

The following list summarizes the contributions of the present work:

1) We propose a mm-wave transmission scheme based on beam selection for B-MIMO, able to achieve near-optimal performances with a reduced RF complexity transceiver, and introduce 3 associated beam selection schemes,
2) We analytically determine the computational complexity of the proposed beam selection schemes, with regards to conventional B-MIMO,
3) We analytically derive the capacity losses caused by beam selection, identifying an upper bound for the proposed techniques,
4) Building on the above, we analyze the performances achieved by the proposed transmission schemes in terms of sum rate and power efficiency, by introducing a joint sum-rate and complexity metric that models the trade-off between performance and complexity.

The paper is organized as follows. In Section II we introduce the channel model used and describe the concepts of beamspace MIMO. Section III presents the selection algorithms, describing in depth their application and implementation. We show complexity analysis studies for all the algorithms proposed in Section IV. Section V is focused on the performance losses caused by beam selection, describing them in an analytical way and obtaining an upper bound for the proposed techniques. Section VI describes the numerical results, while conclusions are discussed in Section VI.

The following notation is used throughout the paper: lower case boldfaced letters denote vectors, upper case boldfaced letters are used for matrices, $tr(\cdot)$ denotes the trace of a matrix, superscripts $(\cdot)^H$ and $(\cdot)^*$ stand respectively for Hermitian transpose and complex conjugate. Matrices are characterized by different subindices and superscripts, here declared and summarized for simplicity: $\mathbf{X}_l$ is used to identify the matrix $\mathbf{X}$ without the l-th row, $\mathbf{x}_{:,l}$ and $\mathbf{x}_{l,:}$ respectively denote the l-th column and l-th row of $\mathbf{X}$.

II. System Model

We consider a downlink communication scenario where the access point (AP), equipped with a discrete lens array and a ZF precoder, communicates with K single-antenna mobile stations or MSs [9], as in Fig. 1. In particular, the DLA at the AP can be analytically modeled as a critically sampled, i.e. $d = \frac{\lambda}{2}$ spaced, uniform linear array (ULA) of length L leading to a signal space dimension n that can be defined analytically as

$$n = \frac{2L}{\lambda}. \tag{1}$$

The parameter n, with DLAs, represents the maximum number of spatial modes that are supported by the antennas, i.e. the total number of orthogonal beams that are supported by the system [17].

For a linearly precoded transmission, the received symbol vector can be expressed as

$$\underline{\mathbf{r}} = \underline{\mathbf{H}}^H \underline{\mathbf{x}} + \underline{\mathbf{n}} \tag{2}$$

where $\underline{\mathbf{H}} = [\underline{\mathbf{h}}_{:,1}, \ldots, \underline{\mathbf{h}}_{:,K}]$ is a $n \times K$ matrix that contains the $n \times 1$ channel vectors $\underline{\mathbf{h}}_{:,1}, .., \underline{\mathbf{h}}_{:,K}$ of all the K MSs, $\underline{\mathbf{x}} = \underline{\mathbf{G}}\underline{\mathbf{s}}$ is the transmitted signal, $\underline{\mathbf{G}}$ is the linear precoding matrix, $\underline{\mathbf{s}}$ is a $K \times 1$ vector that contains all the symbols that have to be transmitted and $\underline{\mathbf{n}}$ is the $K \times 1$ additive white Gaussian noise (AWGN) vector.

It is important to stress that, even though the spatial domain channel model $\underline{\mathbf{H}}$ for DLAs can be analytically modeled as a n sized ULA, the hardware complexity required by these approaches are profoundly different. In fact, as shown in Fig. 1, an AP with DLA requires only a reduced set of RF chains, one for each data stream, and a beam selector to support n narrow beams. A classical MIMO approach instead, where the AP is equipped an equivalent n-dimensional linear array, requires one RF chain for each of the radiating elements, regardless of the number of streams to be transmitted.

A. Channel Model

In order to describe the channel model we need to define the steering-response vectors of the array. In particular for ULA,

we have [16], [35]

$$\mathbf{a}(\theta) = [e^{-j2\pi\theta i}]_{i\in\mathcal{I}(n)} \tag{3}$$

where $\mathbf{a}(\theta)$ is the $n \times 1$ steering vector, $\theta = \frac{d}{\lambda}\sin(\phi)$ is the spatial frequency, whose value ranges between $-0.5 \leq \theta \leq 0.5$ for a critically spaced array, ϕ represents the physical directions $-\pi/2 \leq \phi \leq \pi/2$ and $\mathcal{I}(n) = \{i - (n-1)/2 : i = 0, 1, \ldots, n-1\}$ is a symmetric set of indices centered around 0.

Hence the multipath (MP) channel vector for the k-th MS of a multiuser scenario can be defined as

$$\underline{\mathbf{h}}_{:,k}^{(MP)} = \sum_{i=1}^{N_p} \beta_i^{(k)} \mathbf{a}\left(\theta_i^{(k)}\right) \tag{4}$$

where N_p is the total number of paths, $\beta_i^{(k)}$ and $\theta_i^{(k)}$ are respectively the gain and the direction of the i-th path of the k-th user. In particular, $\theta_i^{(k)}$ can be evaluated in terms of spatial frequencies.

This model however does not take into account the line-of-sight (LoS) component of the propagation, which strongly characterizes the channel in mm-wave frequencies communications. In fact, the small wavelengths λ involved in mm-wave communications lead to narrow and high gain beams with reduced angular spreads. Thanks to these considerations, we can consider the presence of LoS paths for all the MSs without loss of generality.

The channel model for the LoS path can be defined as

$$\underline{\mathbf{h}}_{:,k}^{(LoS)} = \beta_0^{(k)} \mathbf{a}\left(\theta_0^{(k)}\right) \tag{5}$$

where $\theta_0^{(k)}$ represents the direction or position of the k-th user and $\beta_0^{(k)}$ is the complex gain for the LoS path.

Finally, we can define the channel vector for the k-th user as a sum of the two terms (4) and (5), leading to

$$\underline{\mathbf{h}}_{:,k} = \underline{\mathbf{h}}_{:,k}^{(LoS)} + \underline{\mathbf{h}}_{:,k}^{(MP)} = \beta_0^{(k)} \mathbf{a}\left(\theta_0^{(k)}\right) + \sum_{i=1}^{N_p} \beta_i^{(k)} \mathbf{a}\left(\theta_i^{(k)}\right) \tag{6}$$

where the ratio between β_0^2 and $\sum_{i=1}^{N_p} \beta_i^2$ is called Rice factor. Note that the spatial domain model described here implicitly includes the effects of transmit correlation. In fact, since we model the spatial domain for DLA as a critically sampled ULA, different degrees of correlation can be achieved by varying the angle spread [36], [37].

B. Beamspace Representation

The model presented in the previous section can be translated in the beamspace or angular [35] domain via the beamforming matrix $\mathbf{U}$, which represents the operation of a perfectly designed DLA [17]. The beamforming matrix is obtained by computing the steering vectors for n fixed spatial frequencies with uniform spacing [16], [38]. The beamforming matrix is defined analytically as follows

$$\mathbf{U} = \frac{1}{\sqrt{n}} \left[\mathbf{a}(\Delta\theta_0 i)\right]_{i\in\mathcal{I}(n)}, \tag{7}$$

leading to a $n \times n$ matrix where $\Delta\theta_0 = \frac{1}{n}$ is the uniform spacing used.

We can define the (2) in the beamspace domain as follows

$$\mathbf{r} = \mathbf{H}^H \mathbf{x} + \mathbf{n} \tag{8}$$

$$\mathbf{H} = \mathbf{U}\underline{\mathbf{H}} \tag{9}$$

where $\mathbf{x} = \mathbf{U}\underline{\mathbf{G}}\mathbf{s}$, $\mathbf{s} = \underline{\mathbf{s}}$ and $\mathbf{n} = \underline{\mathbf{n}}$ are the transmitted signal, symbol vector and noise in the beamspace domain. The multiplication of the channel matrix $\underline{\mathbf{H}}$ for the beamforming matrix is hence a mapping of the signals for each MS in a new domain of orthogonal beams, defined by $\mathbf{U}$. In the angular or beamspace domain, each row of the channel model $\mathbf{H}$ represents one of the n beams supported by the DLA. The beamforming matrix defined in (7) is unitary, $\mathbf{U}^H\mathbf{U} = \mathbf{U}\mathbf{U}^H = \mathbf{I}$, leading to the following relationships between the spatial and the beamspace domain

$$\underline{\mathbf{x}} = \mathbf{U}^H \mathbf{x}, \quad \mathbf{x} = \mathbf{U}^H \underline{\mathbf{x}}. \tag{10}$$

In addition, the relationship between the channel in the angular-beamspace domain $\mathbf{H}$ and the channel in the spatial domain $\underline{\mathbf{H}}$ is well known. In fact, since the elements of $\mathbf{U}$ are in the form $\frac{1}{\sqrt{n}} e^{-j2\pi kl/n}$, $\mathbf{H}$ is the inverse discrete Fourier transform of the channel matrix in the spatial domain $\underline{\mathbf{H}}$ [16], [35].

III. Beam Selection Techniques

In this paper we apply different selection criteria to identify the beams that will be used by the system during the data transmission. The use of DLA at the transmitter allows us to apply the selection algorithm directly over the channel matrix in the beamspace domain, hence without affecting the beamwidth nor the gain of the antenna pattern. Beam selection can be performed according to different parameters, such as the magnitude of the path [9], the signal-to-interference-ratio (SINR) at the receiver [39], the capacity of the system [40], [41] and minimum error rate [41]. Our main focus is on the application of selection criteria based on the first three characteristics, since our analysis is mostly based on the spectral efficiency of the system.

A. Maximum Magnitude Selection (MM-S)

As a reference to the proposed beam selection schemes, we define as maximum magnitude selection (MM-S) the criterion used in [9]. This criterion takes advantage of the properties of the channel model in the BS domain. The channel matrix $\mathbf{H}$ has in fact a sparse nature, where few elements of the matrix have dominant values near the LoS direction of the MSs. This is a valid assumption for channels where the multipath component of (6) is negligible, but becomes questionable as we introduce additional paths to the model. In order to apply the MM-S we need to define a set of beam indices called sparsity masks. Sparsity masks are used by the AP to identify the dominant beams to be selected for the transmission and are defined as follows

$$\mathcal{M}^{(k)} = \left\{ i \in \mathcal{I}(n) : |h_{i,k}|^2 \geq \xi^{(k)} \max_i |h_{i,k}|^2 \right\} \tag{11}$$

$$\mathcal{M} = \bigcup_{k=1,..,K} \mathcal{M}^{(k)} \tag{12}$$

where $h_{i,k}$ is the i-th element of the k-th column of $\mathbf{H}$, $\mathcal{M}^{(k)}$ is the sparsity mask for the k-th MS and $\xi^{(k)} \in [0, 1]$ is the threshold used to define it. We can see that in order to obtain a minimum number of beams for each of the K MSs, the threshold $\xi^{(k)}$ is chosen independently for each user.

After the sparsity mask, we can define the channel derived from the activation of a subset of beams as

$$\widetilde{\mathbf{H}} = [\mathbf{h}_{l,:}]_{l \in \mathcal{M}} \tag{13}$$

where the sizes $n_d \times K$ of the new channel matrix $\widetilde{\mathbf{H}}$ depend on the number of dominant beams $n_d = |\mathcal{M}|$ identified in the sparsity mask. From (12) we can see that the MM-S algorithm leads to values of n_d which change according to the channel realization. In fact, the user wise selection implemented by MM-S often leads to multiple selections of the same beam for different users and therefore a varying number of required RF chains for different channel realizations and user topologies. As a consequence, a direct application of MM-S in practical systems, where the number of RF chains is fixed, is not viable.

While MM-S selects the strongest channel paths, it can be seen that it is suboptimal in the receive SNR or capacity. Toward this end we propose three selection techniques, described in the following sub-sections.

B. Proposed Maximization of the SINR Selection (MS-S)

In the proposed technique, beams are chosen to maximize the signal to interference ratio at the MS side; we define this selection criterion as Maximization of SINR selection (MS-S). In order to identify the subset of beams used during data transmission, we need to define the SINR metric for our model. The SINR for each user depends on the precoder used at the transmitter, which can described by the precoding matrix $\mathbf{G}$

$$\mathbf{G} = \alpha \mathbf{F} \tag{14}$$

where $\mathbf{F}$ is the precoding matrix without power scaling and α the scaling factor that guarantees $E[\mathbf{x}\mathbf{x}^H] = 1$, defined analytically as

$$\alpha = \sqrt{\frac{\rho}{tr(\mathbf{F}\Lambda_S\mathbf{F}^H)}} \tag{15}$$

where ρ is the signal power and $\Lambda_S = E[\mathbf{s}\mathbf{s}^H]$ is the input covariance matrix, considered unitary and diagonal for our system.

The received SINR of the i-th user is defined as [9]

$$SINR_i(\rho, \mathbf{G}|\mathbf{H}) = \frac{\frac{\rho|\alpha|^2}{K}\left|\mathbf{h}_{:,i}^H\mathbf{f}_{:,i}\right|^2}{\frac{\rho|\alpha|^2}{K}\sum\limits_{m\neq i}\left|\mathbf{h}_{:,m}^H\mathbf{f}_{:,i}\right|^2 + \sigma^2} \tag{16}$$

where $\mathbf{h}_{:,i}^H$ is the Hermitian transpose of the i-th column of $\mathbf{H}$ and σ^2 is the noise power.

In this work, we focus on a practical case where the AP is equipped with a low-complexity zero forcing linear precoder, hence

$$\mathbf{F}_{ZF} = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}. \tag{17}$$

The denominator term in (16) contains two different factors: the first one defines the interference while the second one identifies the noise. Thanks to the properties of ZF precoding we have $\sum_{m\neq i}|\mathbf{h}_{:,m}^H\mathbf{f}_{:,i}|^2 = 0$ and $|\mathbf{h}_{:,i}^H\mathbf{f}_{:,i}|^2 = 1$, leading to the definition of a simplified SINR equation [36]

$$SINR_{i,ZF}(\gamma, \mathbf{G}|\mathbf{H}) = \frac{\gamma|\alpha|^2}{K} \tag{18}$$

where $\gamma = \rho/\sigma^2$ is the signal-to-noise ratio (SNR). The maximization of the SINR can then be obtained simply, by maximizing the scaling factor α.

In order to maximize the SINR, a full search algorithm would compute the SINR for all the possible combination of beam subsets and then choose the subset that leads to the highest value. Such approach leads to an optimal but computationally prohibitive selection because of its $\binom{n}{N}$ possible combinations,[1] where N is the subset size. We propose a suboptimal decremental selection of the beams that identifies the subset of beams with the minimum loss in terms of SINR, shown in Algorithm 1.

Algorithm 1 Incremental MC-S

Input: $\mathbf{H}$
Output: $\widetilde{\mathbf{H}}$

- $\mathbf{C} := \mathbf{H}$
- $\mathbf{F} := \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}$
- **for** $j = 1 \rightarrow n - N$
  - **for** $l = 1 \rightarrow n - j$
    - $\mathbf{F}^{(l)} = \mathbf{C}_l(\mathbf{C}_l^H\mathbf{C}_l)^{-1}$
    - $\alpha^{(l)} = \sqrt{\rho/tr(\mathbf{F}^{(l)}\mathbf{F}^{(l)H})}$
  - **end**
  - $\delta_j = \arg\max_l\{|\alpha^{(l)}|^2\}$
  - $\mathcal{D} = \{\delta_1, \ldots, \delta_j\}$
  - $\mathbf{C} = [\mathbf{h}_{m,:}]_{m\notin\mathcal{D}}$
- **end**
- $\widetilde{\mathbf{H}} = [\mathbf{h}_{m,:}]_{m\notin\mathcal{D}}$

Using (18) we compute the SINR for the reduced system after the elimination of the l-th beam as

$$SINR_{i,ZF}^{(l)}(\gamma, \mathbf{G}|\mathbf{H}_l) = \frac{\gamma\left|\alpha^{(l)}\right|^2}{K} \tag{19}$$

with

$$\alpha^{(l)} = \sqrt{\frac{\rho}{tr\left(\mathbf{F}^{(l)}\mathbf{F}^{(l)H}\right)}} \tag{20}$$

where $\mathbf{H}_l$ represents the channel matrix whose l-th beam has been eliminated, $\mathbf{F}^{(l)}$ is the precoding matrix obtained with

[1] A scenario where $n = 81$ and $N = 40$ leads to $\binom{81}{40} \approx 2 \cdot 10^{23}$ possible subsets, which is computationally prohibitive for a simulation evaluated study. However, previous works on antenna selection for low dimensional systems [42] showed that the performances of decremental approaches are close to the ones achieved by exhaustive search methods.

the $\mathbf{H}_l$ channel model and $\alpha^{(l)}$ is the corresponding scaling factor.[2] Hence we identify the index of the beam to be disabled via the following maximization criterion

$$\delta = \arg\max_{l} \left\{ \frac{\gamma \left|\alpha^{(l)}\right|^2}{K} \right\} \tag{21}$$

where δ is an element of the subset of disabled beams $\mathcal{D}$. Since ρ, K and σ^2 are channel independent we can reduce the maximization criterion to a simple

$$\delta = \arg\max_{l} \left(\left|\alpha^{(l)}\right|^2 \right). \tag{22}$$

In our studies, the selection metric for MS-S derived in (22) is obtained by exploiting the orthogonal properties of ZF precoding. However, the presented technique can be applied independently from the precoding involved at the AP. In fact, following the notation used in (16) and under a generic precoding assumption $\mathbf{G}$, the MS-S algorithm proceeds by maximizing the SINR for the reduced system

$$SINR_l^{(l)} = SINR_i \left(\rho, \mathbf{G}^{(l)} | \mathbf{H}_l \right) \tag{23}$$

where $\mathbf{G}^{(l)}$ represents the precoding matrix that corresponds to the reduced channel model $\mathbf{H}_l$.

C. Maximization of the Capacity Selection (MC-S)

We identify as selections for maximization of the capacity (MC-S) the algorithms whose main objective is the definition of a subset of beams with the minimum loss in capacity from the full system [40]. The MC-S can be performed with two different approaches:

- Decremental-When the algorithm chooses one by one the beams not to be used.
- Incremental-When the algorithm chooses one by one the beams to be used.

It is immediate to see that the difference between the two algorithms resides in the computational costs [40]. The incremental selection is faster when the number of beams to be used in the subset is lower than $n/2$, while the decremental is to be preferred when the number of beams to be used in the subset is higher than $n/2$.

1) Decremental MC-S(DMC-S): The algorithm selects the beams whose elimination causes the minimum loss in terms of capacity. The capacity is defined as follows

$$C(\mathbf{H}) = \log_2 \det(\mathbf{I} + \gamma \mathbf{H}^H \mathbf{H}). \tag{24}$$

If we consider $\mathbf{H}$ as the channel for the full system, we can compute the capacity after the l-th beam has been disabled as [43]

$$C(\mathbf{H}_l) = \log_2 \det\left(\mathbf{I} + \gamma \mathbf{H}_l^H \mathbf{H}_l\right) \tag{25}$$

[2]Note that MS-S does not affect the transmitted power constraint $E[\mathbf{x}\mathbf{x}^H] = 1$. In fact, the system deriving from the selection employs a ZF precoder, which is computed according to the low dimensional channel matrix $\widetilde{\mathbf{H}}$ obtained through MS-S and uses a scaling factor to constrain the average transmit power.

where the channel $\mathbf{H}_l$ is related to the full system matrix according to the following equation.

$$\mathbf{H}_l^H \mathbf{H}_l = \mathbf{H}^H \mathbf{H} - \mathbf{h}_{l,:}^H \mathbf{h}_{l,:}. \tag{26}$$

We can substitute (26) in (25) to show the relationship in terms of capacity between the two channels

$$C(\mathbf{H}_l) = \log_2 \det\left(\mathbf{I} + \gamma \mathbf{H}^H \mathbf{H} - \gamma \mathbf{h}_{l,:}^H \mathbf{h}_{l,:}\right) \tag{27}$$

which can be rearranged to

$$C(\mathbf{H}_l) = \log_2 \det(\mathbf{I} + \gamma \mathbf{H}^H \mathbf{H}) + \log_2 \det\left(\mathbf{I} - (\mathbf{I} + \gamma \mathbf{H}^H \mathbf{H})^{-\frac{1}{2}} \gamma \mathbf{h}_{l,:}^H \mathbf{h}_{l,:} (\mathbf{I} + \gamma \mathbf{H}^H \mathbf{H})^{-\frac{1}{2}}\right) \tag{28}$$

that, thanks to some straightforward algebra, leads to [40]

$$C(\mathbf{H}_l) = C(\mathbf{H}) + \log_2 \left[1 - \gamma \mathbf{h}_{l,:} (\mathbf{I} + \gamma \mathbf{H}^H \mathbf{H})^{-1} \mathbf{h}_{l,:}^H\right]. \tag{29}$$

In particular (29) shows the relationship in terms of capacity between the full system and the system where a beam has been disabled. The algorithm chooses the beam that minimizes the second term on the right-hand side of the relationship, because it is responsible for the capacity loss from the full system. The search criterion can be described analytically as

$$\delta = \arg\min_{l} \left\{ \mathbf{h}_{l,:} (\mathbf{I} + \gamma \mathbf{H}^H \mathbf{H})^{-1} \mathbf{h}_{l,:}^H \right\}. \tag{30}$$

If we consider a fixed number of beams for our system, the algorithm has to compute all the others $n - N$ different beams to eliminate. The above selection is implemented using Algorithm 2 below.

Algorithm 2 Decremental MC-S

Input: $\mathbf{H}, \gamma$
Output: $\widetilde{\mathbf{H}}$

- $\mathbf{K} := \mathbf{H}$
- $\mathbf{B} := (\mathbf{I} + \gamma \mathbf{K}^H \mathbf{K})^{-1}$
- **for** $j = 1 \rightarrow n - N$
  - **for** $j = 1 \rightarrow n - j$
    - $\Omega^{(l)} = \mathbf{k}_{l,:} \mathbf{B} \mathbf{k}_{l,:}^H$
  - **end**
  - $\delta_j = \arg\min_{l} \{\Omega^{(l)}\}$
  - $\mathcal{D} = \{\delta_1, \delta_2, \ldots, \delta_j\}$
  - $\mathbf{B} := \mathbf{B} + \mathbf{B} \mathbf{k}_{\delta_j,:}^H (\gamma^{-1} - \mathbf{k}_{\delta_j,:} \mathbf{B} \mathbf{k}_{\delta_j,:}^H)^{-1} \mathbf{k}_{\delta_j,:} \mathbf{B}$
  - $\mathbf{K} := [\mathbf{h}_{m,:}]_{m \notin \mathcal{D}}$
- **end**
- $\widetilde{\mathbf{H}} = [\mathbf{h}_{m,:}]_{m \notin \mathcal{D}}$

2) Incremental MC-S(IMC-S): The algorithm selects the beams whose contribution in terms of system capacity is the highest. In particular we want to show how the capacity is affected when we add a new row-beam to the channel matrix. By using a similar notation as the previous one for the IMC-S, the

TABLE I
COMPLEXITY IN NUMBER OF OPERATIONS

MM-S	No.	**MS-S**	No.	**DMC-S**	No.	**IMC-S**	No.
$\mathbf{H} \circ \mathbf{H}^*$	$O(Kn)$	$\mathbf{F} = \mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}$	$O(Kn^2)$	$\mathbf{B}$	$O(n^3)$	find	$O(n)$
find	$O(n_b Kn)$	α (n_{del})	$n_{del}O(n^3)$	$\mathbf{k}_{l,:}\mathbf{B}\mathbf{k}_{l,:}^H$ (n_{del})	$n_{del}O(n^2)$	$\mathbf{A}$	$O(n^3)$
		find (n_{del})	$n_{del}O(n)$	find (n_{del})	$n_{del}O(n)$	$\mathbf{A}(N)$	$3NO(n^2)$
		$\mathbf{F}(n_{del})$	$n_{del}O(Kn^2)$	$\mathbf{B}$ (n_{del})	$3n_{del}O(n^2)$	$\mathbf{k}_{l,:}\mathbf{A}\mathbf{k}_{l,:}^H(N)$	$NO(n^2)$
						find (N)	$NO(n)$
Total	$O(Kn)+$ $n_b O(Kn)$	Total	$n_{del}O(n^3)+$ $(1+n_{del})O(Kn^2)$ $+n_{del}O(n)$	Total	$O(n^3)+$ $4n_{del}O(n^2)$ $+n_{del}O(n)$	Total	$O(n^3)+$ $4NO(n^2)+$ $(N+1)O(n)$

capacity of the channel with an additional beam can be defined as [40]

$$C(\bar{\mathbf{H}}, \mathbf{h}_{l,:}) = \log_2 \det\left[\mathbf{I}+\gamma\left(\bar{\mathbf{H}}^H\bar{\mathbf{H}} + \mathbf{h}_{l,:}^H\mathbf{h}_{l,:}\right)\right] \tag{31}$$

where $\bar{\mathbf{H}}$ represents the channel matrix formed by the beams that were previously chosen and $\mathbf{h}_{l,:}$ is the beam we add in the channel model. The (31) can be expressed as a function of the system channel $\bar{\mathbf{H}}$ via the same procedure used for the DMC-S, leading to

$$C(\bar{\mathbf{H}}, \mathbf{h}_{l,:}) = \log_2 \det(\mathbf{I} + \gamma\bar{\mathbf{H}}^H\bar{\mathbf{H}}) + \log_2\left[1 + \gamma\mathbf{h}_{l,:}(\mathbf{I} + \gamma\bar{\mathbf{H}}^H\bar{\mathbf{H}})^{-1}\mathbf{h}_{l,:}^H\right] \tag{32}$$

which has to be maximized by focusing on the second term of the equation with an exhaustive search through all the available beams. In particular

$$\epsilon = \arg\max_{l\notin\mathcal{E}}\left[\mathbf{h}_{l,:}(\gamma^{-1}\mathbf{I} + \bar{\mathbf{H}}^H\bar{\mathbf{H}})^{-1}\mathbf{h}_{l,:}^H\right] \tag{33}$$

where $\mathcal{E}$ represents the subset of enabled beams ϵ. This selection technique is presented analytically in the Algorithm 3, where it uses a recursive update on the matrix inversion, based on the Sherman-Morrison-Woodbury Identity [44].

Algorithm 3 Incremental MC-S

Input: $\mathbf{H}$, γ
Output: $\widetilde{\mathbf{H}}$

- $\mathbf{K} := \mathbf{H}$
- $\mathbf{A} := \gamma\mathbf{I}$
- $\epsilon_1 := \arg\max_l \|\mathbf{k}_{l,:}\|^2$
- **for** $j = 1 \to N-1$
  - $\mathbf{A} := \mathbf{A} - \mathbf{A}\mathbf{k}_{\epsilon_j,:}^H(1 + \mathbf{k}_{\epsilon_j,:}\mathbf{A}\mathbf{k}_{\epsilon_j,:}^H)^{-1}\mathbf{k}_{\epsilon_j,:}\mathbf{A}$
  - **for** $l = 1 \to n-j$
    - $\Omega^{(l)} = \mathbf{k}_{l,:}\mathbf{A}\mathbf{k}_{l,:}^H$
  - **end**
  - $\epsilon_{j+1} = \arg\max_l\{\Omega^{(l)}\}$
  - $\mathcal{E} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_j, \epsilon_{j+1}\}$
- **end**
- $\widetilde{\mathbf{H}} = [\mathbf{h}_{m,:}]_{m\in\mathcal{E}}$

The identity states that for an invertible matrix $\mathbf{A}$ and two or more non-invertible matrices $\mathbf{B}$, $\mathbf{C}$

$$\begin{aligned}(\mathbf{A}+\mathbf{BC})^{-1} &= \left[\mathbf{A}(\mathbf{I}+\mathbf{A}^{-1}\mathbf{BC})\right]^{-1}\\ (\mathbf{A}+\mathbf{BC})^{-1} &= (\mathbf{I}+\mathbf{A}^{-1}\mathbf{BC})^{-1}\mathbf{A}^{-1}\end{aligned} \tag{34}$$

which, thanks to the identity $(\mathbf{I}+\mathbf{P})^{-1} = \mathbf{I} - (\mathbf{I}+\mathbf{P})^{-1}\mathbf{P}$, can be modified to

$$(\mathbf{A}+\mathbf{BC})^{-1} = \left[\mathbf{I} - (\mathbf{I}+\mathbf{A}^{-1}\mathbf{BC})^{-1}\mathbf{A}^{-1}\mathbf{BC}\right]\mathbf{A}^{-1} \tag{35}$$

rearranged with the identity $\mathbf{P}+\mathbf{PQP} = \mathbf{P}(\mathbf{I}+\mathbf{QP})^{-1}$ to

$$(\mathbf{A}+\mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(1+\mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1}. \tag{36}$$

IV. COMPUTATIONAL COMPLEXITY ANALYSIS

This section of the paper analyzes the computational complexity of each of the proposed algorithms, in order to achieve a complete comparison with the approach by [9]. Complexity evaluations for all the algorithms are listed in the Table I, pointing up the number of operations required for each step.

For clarity we emphasize the distinction between the digital signal processing (DSP) complexity, which is the focus of this section, and the RF chain complexity. In fact DSP complexity involves the processor at the transmitter and its impact in power consumption is of the order of $5.76mW/KOps - 22.88mW/KOps$ as for the Virtex family from Xilinx [45], where values are expressed in watts per 10^3 operations. RF complexity, instead, derives by the number of chains used in the transmission. Each chain is characterized by a high number of elements, such as mixer, digital-analogic converter (DAC) and filters, whose values of power consumption are particularly significant. Typical values of power consumption for a single RF chain are of the order of $\sim 30mW$ as in [12], leading to power consumptions in the order of watts, when the amplifier is included in the model.

The first column represents the MM-S criterion as a reference, while the other columns collect the analysis of MS-S, DMC-S and IMC-S respectively. In particular we identify with the notation n_b the number of beams chosen per user by MM-S and with $n_{del} = n - N$ the number of beams to be deactivated in decremental selections. We focus our analysis on the application of the algorithms within a channel realization and on the

operations that dominate the complexity. In order to define the complexity order for each operation, we used the costs presented in the literature [35], [46].

The table shows that the DSP complexity for the MM-S is lower than the other algorithms, as the selection is based only on the amplitude of the paths. As a consequence, the beam selection algorithms presented in the other columns are affected by higher computational complexity. The higher costs are due to the necessity to compute additional elements, such as α for the MS-S or $\mathbf{A}$ and $\mathbf{B}$ for the DMC-S and IMC-S respectively. We decided to keep the constant terms in the complexity computations to show the differences between the DMC-S and IMC-S. Thanks to this notation it is possible to see how DMC-S is more efficient when $n_{del} \leq n/2$ while IMC-S has to be chosen when $N \leq n/2$. Since differences in performances are negligible, the results obtained by these techniques will be addressed as MC-S from now on, without differentiating between incremental or decremental approach.

Finally, it is worth noticing that, even though the MM-S has a lower computational time, it is affected by strong losses in performances in a realistic MP environment, as shown in the results that follow. This consideration makes the presented algorithms relevant in realistic applications, thanks to their appealing trade-off between computational costs and performances.

V. Performance Analysis-Spectral Efficiency Loss

This section is dedicated to the analysis of the spectral efficiency losses caused by the selection of a subset of beams over the full system, providing an analytical study of the performances achieved by the proposed algorithms. The search of the best tradeoff is a critical element of the system design because the selection of a subset of beams benefits from a hardware complexity simplification while suffering a degradation in performances.

The spectral efficiency achievable by a multiuser system can be defined as

$$C = \sum_{i=1}^{K} \log_2(1 + SINR_i). \tag{37}$$

In particular, when using a ZF linear precoder, the detection SINR in a multiuser scenario depends only on the scaling factor α, and is given as [36]

$$SINR_{i,ZF} = \frac{\gamma\rho}{K \cdot tr\left[(\mathbf{H}^H\mathbf{H})^{-1}\right]}. \tag{38}$$

It is immediate to see that (38) leads to the same value for all the users, then the capacity can be evaluated as

$$C = K \log_2(1 + SINR_{ZF}). \tag{39}$$

The losses caused by the elimination of one beam can be defined as the difference between the performances achieved by the full system and by the system when one beam is eliminated

$$\begin{aligned}\Psi^{(l)} &\triangleq K\log_2(1+SINR_{ZF}) - K\log_2\left(1+SINR_{ZF}^{(l)}\right)\\ &= K\log_2\left(1+\frac{\gamma\rho/K}{tr\left[(\mathbf{H}^H\mathbf{H})^{-1}\right]}\right)\\ &\quad - K\log_2\left(1+\frac{\gamma\rho/K}{tr\left[(\mathbf{H}_l^H\mathbf{H}_l)^{-1}\right]}\right)\end{aligned} \tag{40}$$

where $SINR_{ZF}^{(l)}$ represents the SINR for the system without the l-th beam. The equation (40) is particularly useful to show the optimality of the MS-S. In fact, in order to find the best tradeoff between performances and hardware complexity, the losses caused by the selection have to be minimized. The first term in (40) does not depend on the selection because it represents the full system, while the second term depends on the criterion used to identify the l-th beam. Hence the minimum loss $\Psi^{(l)}$ is obtained when the second term is maximized.

In particular, the second term of the equation can be rearranged by using the matrix properties showed in (26) and (36) as

$$K\log_2\left(1+\frac{\gamma\rho/K}{tr\left[\mathbf{R}^{-1}+\mathbf{R}^{-1}\mathbf{h}_{l,:}^H\left(1-\mathbf{h}_{l,:}\mathbf{R}^{-1}\mathbf{h}_{l,:}^H\right)^{-1}\mathbf{h}_{l,:}\mathbf{R}^{-1}\right]}\right) \tag{41}$$

where $\mathbf{R} = \mathbf{H}^H\mathbf{H}$.

Hence, thanks to the properties of logarithms, (40) can be rearranged as

$$\Psi^{(l)} = K\log_2\left(\frac{\left(1+\frac{\gamma\rho/K}{tr(\mathbf{R}^{-1})}\right)}{\left(1+\frac{\gamma\rho/K}{tr\left[\mathbf{R}^{-1}+\mathbf{R}^{-1}\mathbf{h}_{l,:}^H\left(1-\mathbf{h}_{l,:}\mathbf{R}^{-1}\mathbf{h}_{l,:}^H\right)^{-1}\mathbf{h}_{l,:}\mathbf{R}^{-1}\right]}\right)}\right). \tag{42}$$

With some straightforward algebra, (42) can be simplified to the form

$$\Psi^{(l)} = K\log_2(1+\iota) \tag{43}$$

where the parameter ι in the argument of the logarithm of (43) is

$$\iota = \frac{\frac{\gamma}{K}tr(\mathbf{R}\mathbf{h}_{l,:}^H(1-\mathbf{h}_{l,:}\mathbf{S}\mathbf{h}_{l,:}^H)\mathbf{h}_{l,:}\mathbf{S})}{tr(\mathbf{S})^2+\frac{\gamma}{K}tr(\mathbf{S})+tr(\mathbf{S})tr\left(\mathbf{S}\mathbf{h}_{l,:}^H\left(1-\mathbf{h}_{l,:}\mathbf{S}\mathbf{h}_{l,:}^H\right)^{-1}\mathbf{h}_{l,:}\mathbf{S}\right)} \tag{44}$$

where $\mathbf{S} = \mathbf{R}^{-1}$.

Results obtained in (43), (44) can be generalized to identify the global loss caused by the selection of a subset of beams as

$$\Psi = K\log_2\left(1+\frac{\frac{\gamma}{K}tr(\mathbf{T}_{\mathcal{D}})}{tr\left(\mathbf{R}_{\mathcal{E}}^{-1}\right)^2+\frac{\gamma}{K}tr\left(\mathbf{R}_{\mathcal{E}}^{-1}\right)+tr\left(\mathbf{R}_{\mathcal{E}}^{-1}\right)tr(\mathbf{T}_{\mathcal{D}})}\right) \tag{45}$$

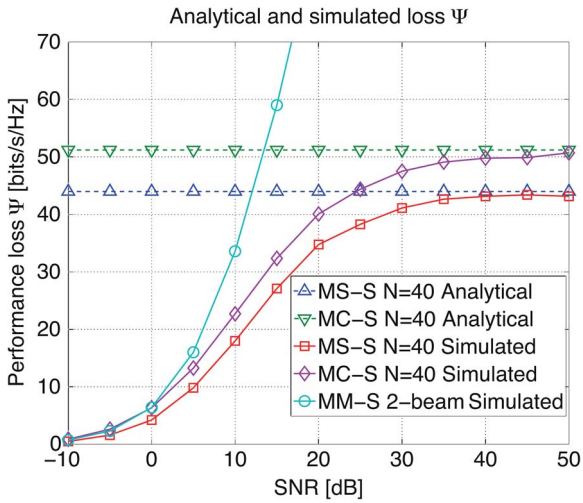

Fig. 2. Comparison analytical and simulated Ψ for a $n = 81$ and $K = 40$ system.

where $\mathbf{R}_{\mathcal{E}} = \mathbf{H}_{\mathcal{E}}^H \mathbf{H}_{\mathcal{E}}$, $\mathbf{T}_{\mathcal{D}} = \mathbf{R}_{\mathcal{E}} \mathbf{H}_{\mathcal{D}}^H (\mathbf{I} - \mathbf{H}_{\mathcal{D}} \mathbf{R}_{\mathcal{E}}^{-1} \mathbf{H}_{\mathcal{D}}^H)^{-1} \mathbf{H}_{\mathcal{D}} \mathbf{R}_{\mathcal{E}}^{-1}$ and the subindices $\mathcal{E}$ and $\mathcal{D}$ represent the enabled and disabled subset of beams respectively. In particular $\mathbf{H}_{\mathcal{D}} = [\mathbf{h}_{m,:}]_{m \in \mathcal{D}}$ and $\mathbf{H}_{\mathcal{E}} = [\mathbf{h}_{m,:}]_{m \in \mathcal{E}}$.

Hence the loss Ψ is a function of γ and approaches an upper bound [42] as $\gamma \to \infty$, defined in the following equation

$$\Psi \leq K \log_2 \left(1 + \frac{tr(\mathbf{T}_{\mathcal{D}})}{tr\left(\mathbf{R}_{\mathcal{E}}^{-1}\right)} \right). \tag{46}$$

The analytical results of the loss are confirmed by simulations ($n = 81$ and $K = 40$ system), as shown in Fig. 2 for a MP scenario. In particular, the upperbounds derived through (46) are indicated in the legend as *MC-S N = 40 Analytical* and *MS-S N = 40 Analytical* for the selection over capacity and SINR respectively, while the numerical values are addressed as *Simulated* for all the selection techniques.

The MM-S criterion is characterized by fluctuations in the size of the beam subset, leading to losses that are not limited by the upperbound defined in (46).

VI. Numerical Results

In this section we present the numerical results obtained through Monte Carlo simulations over 10000 channel realizations. For comparison to [9] we assume a transmission scheme where the AP is equipped with an ULA of $n = 81$ elements and communicates with $K = 40$ MSs, equipped with a single isotropic antenna each.

We define two different channel scenarios both with perfect channel state information at the transmitter[3]: one with only the LOS component (5) in accordance to [9] and one with the additional MP components as in (6) with $N_P = 2$, in line with [9]. We define analytically the complex path gains as

$$\beta_i = |\beta_i| e^{-j\psi_i} \tag{47}$$

where:

- MP component ($i \neq 0$): $|\beta_i|^2 = -10dB$ and ψ_i is uniformly distributed between 0 and 2π
- LoS component ($i = 0$): $|\beta_0|^2 = 0dB$ and $\psi_0 = 0$.

In order to have a simplified definition for the angles of arrival in the MP component, we consider a scenario where the distance between AP and MSs is wide enough so that $|\theta_i|$ is uniformly distributed between $\left[\Delta\theta - \frac{\Delta\theta}{4}, \Delta\theta + \frac{\Delta\theta}{4}\right]$ and $sign(\theta)$ is chosen randomly.

We applied the algorithms with two different approaches: one where we fix the total number of beams to be used and one where we choose the beams in order to capture a certain percentage η of the total channel power σ_c^2. Accordingly, in the figures, we use the following notation: *Full System* to denote the performances obtained by the scheme without beam selection, *MM-S 2-beam* to identify the performances obtained by magnitude selection with $n_b = 2$ in accordance to [9], $\eta = 95\%$ to classify the approach where the percentage of channel power captured by the subset of beams is fixed, $N = 40$ and $N = K$ to address the approach where the maximum number of beams is fixed at 40 or at the number of users K respectively.

It is worth to notice that in practical systems the number of RF chains is generally fixed, making the MM-S of [9] inapplicable and the $N = \{40, K\}$ approaches shown here particularly relevant in realistic scenarios.

For the $\eta = 95\%$ approach,we used the following definition of channel power

$$\sigma_c^2 = tr(\mathbf{H}\mathbf{H}^H). \tag{48}$$

Hence we can define the captured channel power ratio η as

$$\eta = \frac{tr(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H)}{\sigma_c^2}. \tag{49}$$

A. Spectral Efficiency

In this section we show the spectral efficiencies in B-MIMO achieved with different selection algorithms when the transmitter is equipped with a ZF precoder. In particular we compare our results with the performances we achieve: a) with a full rank system and b) with the selection criterion showed in [9].

If we use the previous definition of SINR in (16) we can compute the spectral efficiency as

$$C(SINR, \mathbf{G}|\mathbf{H}) = \sum_{i=1}^{K} \log_2 \left(1 + SINR_i(\gamma, \mathbf{G}|\mathbf{H})\right) \tag{50}$$

The presented formula is defined for the full channel model $\mathbf{H}$, but can be applied directly for the low dimensional channel after beam selection, by replacing $\mathbf{H}$ with $\widetilde{\mathbf{H}}$.

In Figs. 3 and 4 we can see the spectral efficiency as a function of the SNR for both approaches, $N = 40$ and $\eta = 95\%$ in the legend, and both proposed algorithms, MS-S and MC-S in the legend. Since both incremental and decremental MC-S

[3]Perfect CSI is a common assumption in the literature [13] for systems that involve AS at the transmitter. Due to the sizes of the systems involved in mm-wave communications, the acquisition of channel state information represents a critical step. Recent works on M-MIMO approached the problem, with the aim to reduce the signal processing complexity [47] or the time [48] required for CSI acquisition. In our scenario, to retain the benefit of reduced RF-chain operation, a trivial approach for the CSI provisioning would be by scheduled acquisition and RF switching, although more sophisticated and efficient approaches can be found in the literature.

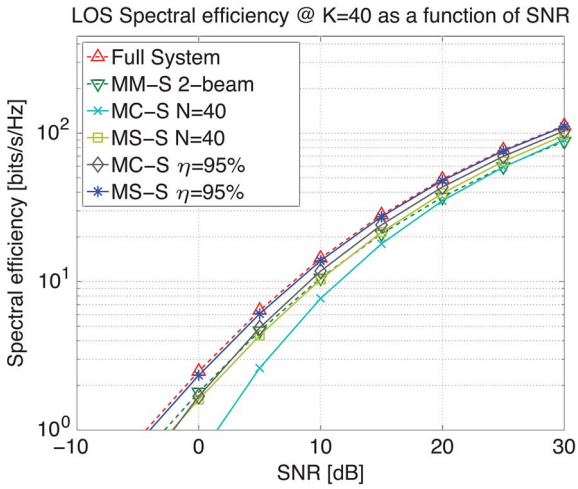

Fig. 3. Spectral efficiency as a function of SNR[dB] in a Line of Sight scenario.

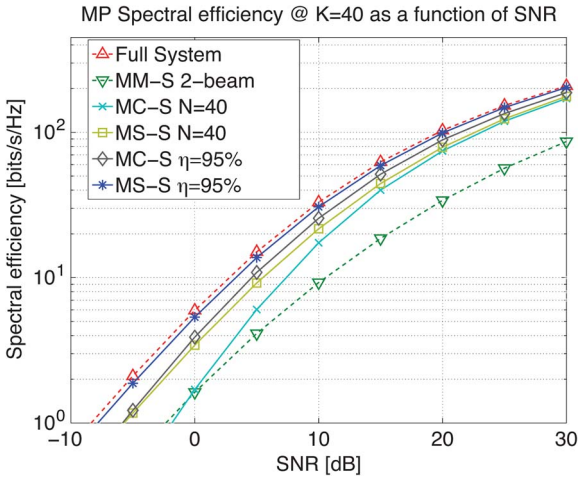

Fig. 4. Spectral efficiency as a function of SNR[dB] in a Multipath scenario with $N_P = 2$.

achieve the same performances, DMC-S has been used to obtain the results showed, because it introduces complexity savings, as explained above. When comparing the performances of MC-S and MS-S, we can see a gap in the low SNR region. This gap, beneficial for MS-S, is justified by the different metrics used by the two algorithms. In fact, whilst MC-S does not consider the precoding involved at the AP, the MS-S algorithm maximizes the SINR at the receiver for the particular ZF precoding used here, by maximizing the scaling factor α. The impact of this difference over the received SINR is described analytically in (16), where the noise component of the denominator is inversely proportional to the scaling factor.[4]

The losses caused by beam selection in $N_p = 0$ scenarios are almost negligible, as we can see in Fig. 3. In particular the figure shows that the $N = 40$ approach is characterized by higher performance degradations for the low-SNR region, but it starts to achieve similar or better performances than the MM-S for

[4]Hence, the gap between the two techniques is wider in the low-SNR region, since the dependence of the SINR over the scaling factor is more visible at high values of noise power.

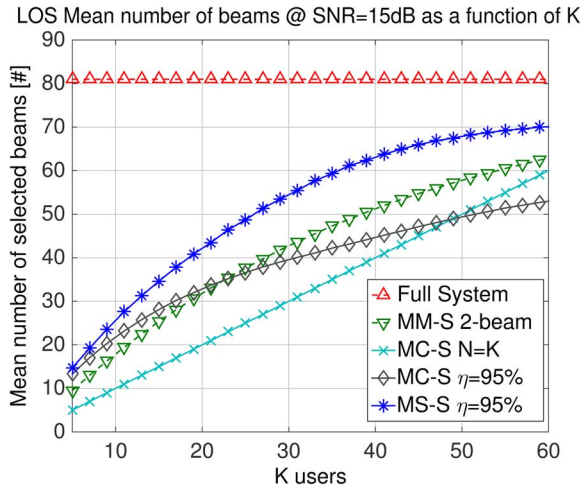

Fig. 5. Mean number of beams (RF chains) used for transmission as a function of the number of users in a Line of Sight scenario.

$SNR \geq 10dB$ for both MC-S and MS-S. If we apply the beam selection according to the $\eta = 95\%$ approach, then performances increase greatly for both algorithms, arriving to almost optimal performances with the MS-S.

Fig. 4 shows the spectral efficiencies obtained in the MP environment. For the $N = 40$ study we see how the MC-S algorithm performs well as we increase the SNR, while the MS-S algorithm outperforms the previous MM-S approach even in low SNR regions. When we set the number of beams according to the $\eta = 95\%$ approach instead, we can see how the capacities obtained by both algorithms are very close to the optimal performances. In particular the MS-S algorithm performs very closely to the full-system with a negligile degradation.

It is important to notice that the improvement in performances between the $N = 40$ and the $\eta = 95\%$ approach resides in a higher number of selected beams, as shown in the following section where we focus on this aspect.

B. Mean Number of Beams

In this section we show the mean number of selected beams for both algorithms as a function of the number of users in the system. The beam usage is a fundamental parameter for our study, because it provides an immediate evaluation of the RF complexity reduction achieved by beam selection at the transmitter. In particular, the $N = K$ environment is characterized by a number of beams that is a linear function of K because of the convention we used. This holds for both MC-S and MS-S and leads to matching results, presented in Figs. 5 and 6 with MC-S $N = K$ only.

In Fig. 5 we can see how the MM-S *2-beam* selects a number of beams that grows constantly and rapidly with the number of users in the scenario. In particular we can see that the $N = K$ approach leads to an higher simplification of the transmitter than the other approaches when $20 \leq K \leq 46$. When we apply the selection with the $\eta = 95\%$ approach, the subset defined by the MC-S is characterized by a number of beams which is always lower than the MM-S, except for $K = 20$. On the other

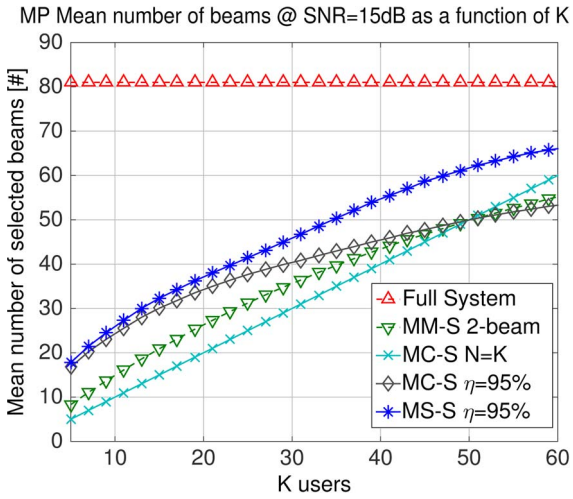

Fig. 6. Mean number of beams (RF chains) used for transmission as a function of the number of users in a Multipath scenario with $N_p = 2$.

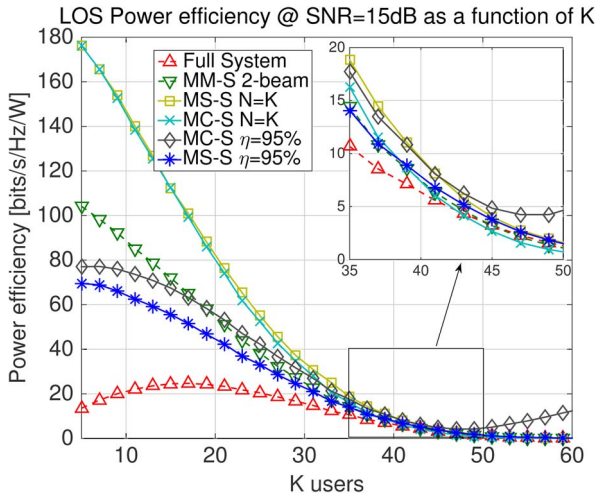

Fig. 7. Power efficiency as a function of K in a Line of Sight scenario with $P_t = 15dBm$.

hand, the MS-S algorithm is affected by a wider selection of beams than the MM-S, while still providing an interesting simplification gain compared to the full system.

In a more realistic multipath scenario, see Fig. 6, the MM-S technique uses a higher number of beams than the $N = K$ approach with both MC-S and MS-S until we keep the number of users $K \leq 50$. In particular, it is interesting to notice how this simplification benefit is accomplished while providing at the same time higher spectral efficiencies. When we apply the MC-S with a $\eta = 95\%$ approach the gains in terms of number of beams are visible for systems where $K \geq 50$, while the MS-S uses always bigger subsets than the MM-S. For the case with 40 users, we can see that the MC-S selects in average only $\sim$2 more beams than the MM-S and yet provide a considerable benefit in terms of spectral efficiency, leading to a very advantageous trade-off. In a system with $K = 40$, the MS-S algorithm in a MP scenario is affected by a selection of a wider number of beams, but at the same time is able to provide near-optimal performances with an significant complexity simplification compared to the full system.

In both Figs. 5 and 6 we consider a $SNR{=}15dB$ scenario, however, the effects of this assumption over the number of selected beams are negligible. In fact, while selections with a fixed number of beams $N = K$, together with MM-S and MS-S are independent from the SNR, the channel power approach with MC-S showed imperceptible differences in a low SNR scenario.

C. Power Efficiency

In order to provide an evaluation of the trade-off between performance and RF complexity in the practical implementation in terms of RF chains required, we show the power efficiency values obtained by the different selection algorithms we used. We follow the definition of transmit power efficiency used in [22] based on the modelling [49]

$$\varepsilon_P = \frac{C}{P_t + N \cdot P_{RF}} \tag{51}$$

where C represents the spectral efficiency in $[bits/s/Hz]$, P_t the transmitted power of the system in $[Watt]$, N the number of transmitting beams and P_{RF} the power consumed in the components per RF chain in $[Watt]$. We use practical values for $P_{RF} = 34.4mW$ [22], accounting for mixer, DAC and filters, and $P_t = 15dBm$ [4], to model a small cell transmission. This metric is particularly useful to show the effects of the selection of a reduced number of beams on the power needed by the system, together with the effects on the average spectral efficiency of the system.

Fig. 7 illustrates how the $N = K$ approach, with both the MC-S and MS-S, outperforms greatly all the others until the number of users $K \leq 31$. This is due to the fact that the number of beams, and hence RF chains, we use is much smaller than the other approaches, yielding a great reduction in power consumption. The $\eta = 95\%$ approach is characterized by lower efficiencies than with $N = K$ for reduced number of users, due to the independence of the selection criterion from the number of users, but once $K \geq 35$ the MC-S $\eta = 95\%$ approach keeps having higher performances than all the other techniques. The channel power approach over MC-S shows an interesting behavior in the highly populated scenario, i.e. $K > 50$, where the values of power efficiency start to increase. This behavior is explained by Fig. 5, where the mean number of beams selected by MC-S $\eta = 95\%$ when $K > 50$ is lower than all the other approaches and even lower than the number of users in the scenario. Consequently, the increasing spectral efficiency, combined with lower power requirements, leads to an increase in terms of power efficiency. Lastly, when we apply the MS-S we can see that the $\eta = 95\%$ approach, in $K \geq 36$ scenarios, is characterized by higher values than the MM-S algorithm. This is due to the fact that the effects of a higher number of selected beams are mitigated by near optimal spectral efficiency values.

In Fig. 8 we can see that the approach with $N = K$ is still the best one for system with reduced number of users. In particular we can see that the two approaches lead to similar high values of power efficiency in the low populated scenario, but MC-S is gradually outperformed by the MS-S algorithm as we increase

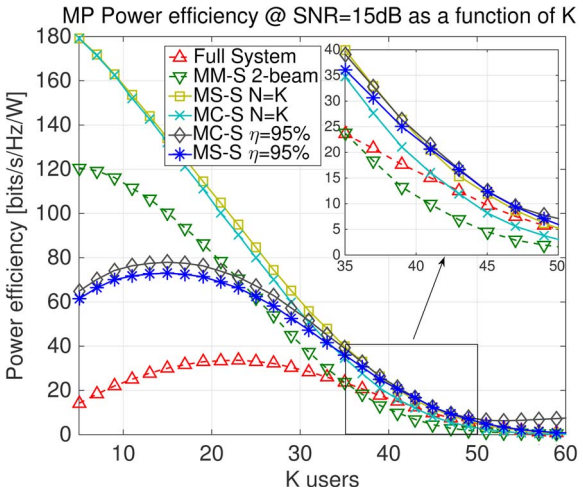

Fig. 8. Power efficiency as a function of K in a Multipath scenario with $N_p = 2$ and $P_t = 15dBm$.

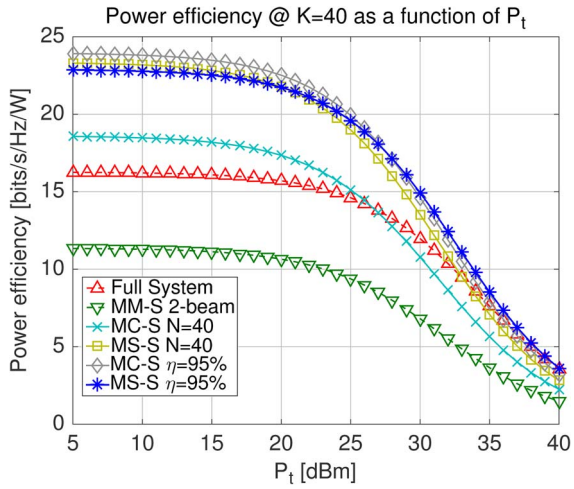

Fig. 9. Power efficiency as a function of P_t.

the number of users. This is due to the fact that the interference among users, which is optimized by MS-S, becomes more relevant as the scenario gets more populated. When we apply the algorithms with the $\eta = 95\%$ approach we can see that the MC-S technique starts outperforming all the others after the number of users increases to $K \approx 35$. This can be easily explained by the high values of spectral efficiency this transmission scheme leads to.

In the end it is worth noticing how the values of power efficiency obtained by the MM-S criterion rapidly decrease with the number of users in the scenario, because of the lower capacities obtained with such approach when the effects of scattering and multipath increase.

In Figs. 7 and 8 we focused our attention on high SNR regime, in a $SNR = 15dB$ scenario. However, from Figs. 3 and 4, we can infer that the benefits of the proposed techniques extend also to low SNR scenarios, as the performance gap in terms of spectral efficiency in comparison with the full system is still narrow and the hardware complexity savings are not affected by the *SNR*. Moreover, as shown in Fig. 2, the losses of the proposed techniques in comparison with the full system decrease together with the SNR. As a consequence, this suggests that the performance trends in terms of power efficiency are not affected by the SNR.

In Fig. 9 we can see the dependence of the efficiency on the transmitted power. In particular, we use a scenario where $K = 40$ and $SNR = 15dB$. As in the previous figures, we can see that the $\eta = 95\%$ approach with MC-S outperforms all the others. As we would expect, performances decrease with higher values of transmitted power, because the RF-chain term in (51) becomes less relevant. In the same figure it is visible how the MM-S technique performs poorly compared to the others because of the low values of spectral efficiency and higher number of beams used. The $N = 40$ strategy provides better performances than the full system when $P_t \leq 26dBm$ for MC-S, and when $P_t \leq 34dBm$ for MS-S. Even though with a fixed number of beams N both the capacity and the SINR selection perform better than the system with no selection, it is interesting to notice how the MS-S is characterized by an additional $\sim 6[bits/s/Hz/W]$ gain.

VII. Conclusion

In this paper, we have introduced several beam selection techniques for B-MIMO that allow to reduce the RF complexity of mm-wave transmitters while obtaining near-optimal performances, in both line of sight and multipath environments. The transmission schemes showed in our studies are particularly interesting for mm-wave systems because of their characteristic high-dimensionality which makes the RF chain costs and power consumption important.

In particular, our analytical and simulation results prove that with beam selection algorithms it is possible to achieve higher power efficiencies than a full system while reducing the transceiver RF complexity according to the number of MSs. We also demonstrated that beam selection algorithms with a channel power approach can lead to near-optimal performances in both a LoS and MP scenario, while still achieving significant RF complexity reductions.

References

[1] "Cisco visual networking index: Global mobile data traffic forecast update 2012–2017," Cisco, San Jose, CA, USA, White_paper_c11-520862, 2013.

[2] B. Bosco, S. Franson, R. Emrick, S. Rockwell, and J. Holmes, "A 60 GHz transceiver with multi-gigabit data rate capability," in *Proc. IEEE Radio Wireless Conf.*, Sep. 2004, pp. 135–138.

[3] J. Wells, "Faster than fiber: The future of multi-G/s wireless," *IEEE Microw. Mag.*, vol. 10, no. 3, pp. 104–112, May 2009.

[4] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-ghz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436 Aug. 2011.

[5] T. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[6] "Millimeter Wave Propagation: Spectrum Management Implications," Fed. Commun. Commiss., Office Eng. Technol., Washington, DC, USA, Bulletin Number 70, 1997.

[7] "The use of the radio frequency spectrum above 30 GHz: A consultative document," Radiocommun. Div., Dept. Trade Ind., London, ON, Canada, 1988.

[8] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[9] A. M. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *Proc. IEEE GLOBECOM*, 2013, pp. 3784–3789.

[10] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[11] A. Hakkarainen, J. Werner, K. R. Dandekar, and M. Valkama, "Widely-linear beamforming and RF impairment suppression in massive antenna arrays," *J. Commun. Netw.*, vol. 15, no. 4, pp. 383–397, Aug. 2013.
[12] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.
[13] S. Sanayei and A. Nosratinia, "Antenna selection in mimo systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.
[14] S. Sanayei and A. Nosratinia, "Capacity of MIMO channels with antenna selection," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4356–4362, Nov. 2007.
[15] A. Molisch and M. Win, "MIMO systems with antenna selection," *IEEE Microw. Mag.*, vol. 5, no. 1, pp. 46–56, Mar. 2004.
[16] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
[17] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.
[18] A. Sayeed and N. Behdad, "Continuous aperture phased MIMO: A new architecture for optimum line-of-sight links," in *Proc. IEEE APSURSI*, Jul. 2011, pp. 293–296.
[19] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.
[20] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, Jul. 2004.
[21] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Vector perturbation based on symbol scaling for limited feedback MIMO downlinks," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 562–571, Feb. 2014.
[22] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Computationally efficient vector perturbation using thresholded optimization," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1880–1890, May 2013.
[23] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Maximizing Energy-Efficiency in the Vector Precoded MU-MISO Downlink by Selective Perturbation," *IEEE Trans. Wireless Comms.*, vol. 13, no. 9, pp. 4974–4984, Sep. 2014.
[24] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Interference optimization for transmit power reduction in Tomlinson-Harashima precoded MIMO downlinks," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2470–2841, May 2012.
[25] A. Garcia and C. Masouros, "Power-Efficient Tomlinson-Harashima Precoding for the Downlink of Multi-User MISO Systems," *IEEE Trans. Comms.*, vol. 62, no. 6, pp. 1884–1896, Jun. 2014.
[26] A. Garcia and C. Masouros, "Power Loss Reduction for MMSE-THP with Multidimensional Symbol Scaling," *IEEE Comms. Let.*, vol. 18, no. 7, pp. 1147–1150, Jul. 2014.
[27] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 86–105, Sep. 2007.
[28] C. Masouros and E. Alsusa, "Two-Stage transmitter precoding based on data-driven code hopping and partial zero forcing beamforming for MC-DCMA communications," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3634–3645, Jul. 2009.
[29] C. Masouros, "Correlation rotation linear precoding for MIMO broadcast communications," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 252–262, Jan. 2011.
[30] C. Masouros and E. Alsusa, "Soft transmitter precoding for the downlink of DS/CDMA communication systems," *IEEE Trans. Veh. Tech.*, vol. 59, no. 1, pp. 203–215, Jan. 2010.
[31] C. Masouros, T. Ratnarajah, M. Sellathurai, C. Papadias, and A. Shukla, "Known Interference in Wireless Communications: A Limiting factor or a Potential Source of Green Signal Power?," *IEEE Comms. Mag.*, vol. 51, no. 10, pp. 162–171, Oct. 2013.
[32] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.
[33] T. Haustein, C. Von Helmolt, E. Jorswieck, V. Jungnickel, and V. Pohl, "Performance of mimo systems with channel inversion," in *Proc. IEEE VTC*, 2002, vol. 1, pp. 35–39.
[34] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
[35] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K., Cambridge Univ. Press, 2005.
[36] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Large-Scale MIMO transmitters in fixed physical spaces: The effect of transmit correlation and mutual coupling," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2794–2804, Jul. 2013.
[37] C. Masouros, and M. Matthaiou, "Physically Constrained Massive MIMO: Hitting the Wall of Favorable Propagation," *IEEE Comms. Let.*, vol. 19, no. 5, pp. 771–774, May 2015.
[38] A. S. Y. Poon, R. W. Brodersen, and D. Tse, "Degrees of freedom in multiple-antenna channels: A signal space approach," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 523–536, Feb. 2005.
[39] M. Sadek, A. Tarighat, and A. H. Sayed, "Active antenna selection in multiuser MIMO communications," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1498–1510, Apr. 2007.
[40] A. Gorokhov, D. Gore, and A. Paulraj, "Receive antenna selection for MIMO flat-fading channels: theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2687–206, Oct. 2003.
[41] I. Berenguer, X. Wang, and V. Krishnamurthy, "Adaptive MIMO antenna selection via discrete stochastic optimization," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4315–4329, Nov. 2005.
[42] P.-H. Lin and S.-H. Tsai, "Performance analysis and algorithm designs for transmit antenna selection in linearly precoded multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 4, pp. 1698–1708, May 2012.
[43] A. Gorokhov, "Antenna selection algorithms for MEA transmission systems," in *Proc. IEEE ICASSP*, May 2002, vol. 3, pp. 2857–2860.
[44] H. Henderson and S. Searle, "On deriving the inverse of a sum of matrices," *Siam Rev.*, vol. 23, no. 1, pp. 53–60, Jan. 1981.
[45] D. Curd, "Power consumption in 65 nm FPGAs," Xilinx, San Jose, CA, USA, White Paper, Feb. 2007.
[46] R. Hunger, "Floating point operations in matrix-vector calculus," Technische Universität München, München, Germany, Oct. 2005.
[47] H. Prabhu, J. Rodrigues, O. Edfors, and F. Rusek, "Approximative matrix inverse computations for very-large mimo and applications to linear pre-coding systems," in *Proc. IEEE WCNC*, Apr. 2013, pp. 2710–2715.
[48] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser mimo systems," in *Proc. IEEE WCNC*, Apr. 2013, pp. 2890–2895.
[49] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

Pierluigi V. Amadori (S'14) received the M.Sc. degree with honours in telecommunications engineering from the University of Rome La Sapienza, Rome, Italy, in 2013. Between 2012 and 2013, he held a JPL Visiting Student Researchers Program position in the Jet Propulsion Laboratory Pasadena, CA, USA. He is currently pursuing the Ph.D. degree in the Department of Electrical & Electronic Engineering at University College London, London, U.K. His main research interests include wireless communications with emphasis on large antenna array systems and energy efficient communications.

Christos Masouros (M'06–SM'14) received the Diploma in electrical and computer engineering from the University of Patras, Greece, in 2004, the M.Sc. degree by research and the Ph.D. degree in electrical and electronic engineering from the University of Manchester, U.K., in 2006 and 2009, respectively. He is currently a Lecturer in the Department of Electrical & Electronic Engineering, University College London. He has previously held a Research Associate position in University of Manchester, and a Research Fellow position in Queen's University Belfast, U.K.

He holds a Royal Academy of Engineering Research Fellowship, 2011-2016, and is the Principal Investigator of the EPSRC project EP/M014150/1 on large scale antenna systems. His research interests lie in the field of wireless communications and signal processing with particular focus on green communications, large scale antenna systems, cognitive radio, interference mitigation techniques for MIMO and multicarrier communications.