Set Optimization for Efficient Interference Alignment in Heterogeneous Networks

Daniel Castanheira, Adão Silva, Atílio Gameiro
DETI, Instituto de Telecomunicações, University of Aveiro,
Emails: dcastanheira@av.it.pt, asilva@av.it.pt, amg@ua.pt

Abstract—To increase capacity and offload traffic from the current macro-cell cellular system operators are considering the deployment of small-cells. It is expected that both the small and macro-cells will coexist in the same spectrum resulting in unsustainable levels of interference. Interference alignment is considered as an effective method to deal with such interference. By using interference alignment the small-cells align their transmission along a common direction to allow the macro-cell receiver to completely remove it. It is clear that if the two systems have no limitations on the information that may be exchanged between them to perform the signal design, then the performance may be improved in comparison to the case of no or partial cooperation. However, this full cooperation strategy requires a high-rate connection between the macro and small-cells, which may not be available. To overcome this problem we consider that the alignment direction is selected from a finite set, known to both macro and small-cell terminals. We provide sufficient conditions for this set that guarantee full-diversity, at the macro-cell, and propose an efficient method to optimize the set elements. Results show that an alignment set with a description length of 1 bit is enough to achieve the same diversity as in the case where an infinite amount of information is exchanged between both systems. The proposed set optimization method achieves better performance than random vector quantization and similar performance to Grassmannian quantization.

Index Terms—Small-cells, Interference Alignment, Zero-Forcing, MIMO Systems, Diversity Methods, Codebook Design, Rayleigh Channels, Feedback, Random Vector Quantization

I. INTRODUCTION

Small-cells are being considered by the operators as a solution to overcome the capacity limitations of the current macro-cell cellular system [1], [2]. Beyond the capacity improvements, small-cells offer other advantages for the operators. Namely, they are easier to deploy, less expensive and more energy efficient, just to mention a few benefits. Nevertheless, due to the expected extensive deployment of small-cells and costs involved in the acquisition of new spectrum licenses the two systems should reuse the same spectrum as much as possible. The macro-cell system is the owner of the spectrum license and in the literature of cognitive radio (CR) the macro-cell terminals are denominated as primary users. Similarly, the small-cell terminals use the spectrum of the former in an opportunistic way and are denominated secondary. If not carefully designed the secondary signal may cause harmful interference on the primary [3]. To deal with the interference problem the secondary terminals must sense the environment looking for unused resources, in the primary network, which can be used for its own transmission. These opportunities can be in time, frequency or even in space.

The primary terminals may not use all the spatial dimensions because of physical constraints that do not allow a number of antennas as large as in the BS, or as pointed out in [4] where optimal power allocation may lead to some unused dimensions. In both cases this leads to free spatial dimensions that can be exploited by the secondary network. In fact recent research points out that active use of the spatial domain through multiple antenna techniques is of utmost importance to minimize the interference level in a heterogeneous network environment. Multiple antenna techniques in combination with CR have been considered in several publications (see e.g. [5] [6]). A new research area is to apply beamforming in a CR manner so that in addition to spectrum sensing also direction of arrival information is exploited [7] [8]. This adds a new dimension for CR systems and enables reuse of the frequencies in spatial domain. In [7] the authors proposed the use of the angle dimension as a new spectrum opportunity, for secondary transmission. A signal to interference plus noise (SINR) balancing technique for the downlink of a cognitive radio network has been proposed in [8]. The authors propose a beamforming technique to maximize the worst secondary user SINR, while ensuring that the interference leakage to the primary is below specific thresholds. In [9] the authors proposed to maximize the throughput of the secondary user under the constraint that the interference to primary receivers is below a certain threshold.

Recently, in [10], [11] and [12] the authors have proposed the use of interference alignment (IA) [13], [14] to mitigate the interference of small-cell user terminals (UTs) towards the macro-cell BS. In [10] the authors proposed a new interference alignment scheme that successively creates transmit beamforming vectors for the small-cell terminals and for the macro BS assuming that they have different number of transmit antennas. A scheme that takes advantage of the OFDM cyclic prefix for interference alignment is proposed in [11]. The work in [12] studied several IA techniques with different levels of inter-system information sharing: a coordinated IA with large information sharing requirements, a static IA and also an uncoordinated IA with no need for information sharing. IA is a precoding technique that is able to achieve the maximum degrees of freedom of the interference channel, in a variety of settings [13]. For some cases (e.g. constant channels) this may require long symbol extensions either across time or frequency [15]. IA works by dividing the receiver space in two parts, one for interference and the other for the intended signal. By using IA all inter-user interference is aligned in the interfer-

ence space, restricting in that way the interference subspace dimensions. Closed form solutions for IA are only available for some specific cases [16]. For the other cases iterative methods may be used [16], [17], [18]. In [16] two iterative algorithms that utilize the reciprocity of wireless networks to achieve interference alignment with only local channel knowledge at each node are proposed. A convergent alternating minimization approach to IA has been proposed in [17]. In [18] the authors presented two algorithms that optimize the precoding subspaces by maximizing the data rate performance while maintaining the achievable degrees of freedom.

A key point in cooperative based systems is the amount of information that needs to be exchanged between the cooperating nodes. Full cooperation between all entities allows to achieve optimum performance. On the other hand, if no cooperation is considered the system performance is reduced to a minimum. The first case is optimum in terms of performance, but requires a huge amount of inter-system information sharing. The pursuit of schemes that achieve close to optimum performance but have reduced information exchange is thus of paramount importance for practical applications.

The topic of limited information IA has been addressed in several publications. By considering that the channels may be accurately estimated at the receiver end, the channel state information (CSI) may be fed back from the receivers to the transmitters. This information can then be used to design the IA precoders. By quantizing the channel over the composite Grassmann manifold the authors, in [19] [20], showed that for a MIMO channel, IA achieves the full multiplexing gain as long as the feedback bit rate scales sufficiently fast with the SNR. In [21] the authors proposed the use of analog feedback. Multiplexing gain is preserved if the feedback link quality is comparable to the one of the forward channel. In [22] an IA scheme based on Random Vector Quantization (RVQ) for the constant MIMO interference channel is proposed. The previous strategies neglect the temporal correlation in the channel. In [23] the authors proposed a Grassmannian differential feedback method to reduce feedback overhead by exploiting both the channels temporal correlation and Grassmannian structure.

Contrarily to the previously proposed schemes, in this manuscript we consider that the terminals participating in the alignment process have different priorities, i.e there are primary and secondary terminals. In other words, we may say that the primary defines a set of unused space dimensions, i.e. a subspace that is unused by the primary network and where transmission of secondary terminals will not cause any interference. These unused space dimensions are what we call the alignment directions. The small-cells align all the inter-system interference along a common direction (the alignment direction), so that no interference is generated on the primary network. The selection of the optimum direction, from the point of view of the primary network, requires full-cooperation between the two systems and thus heavy exchange of information. To limit the amount of information sharing we consider that the alignment direction is selected from a finite set (codebook), at the primary BS. The main contributions of this work are the following

- Extension of the IA technique, proposed in [12], for a generic macro-cell antenna configuration.
- Analysis of the extended scheme, when only limited information exchange between the two systems is allowed.
- As a result of the previous analysis sufficient conditions on the alignment set are provided so that the diversity order achieved at the macro-cell link, is the same as in the case where the set dimensions tend to to infinity.
- Proposal of an efficient method to optimize the elements of the alignment set for a given set cardinality.

This paper is organized as follows: Section II introduces the system model. Section III presents the IA precoders and filtering matrix. In section IV we provide sufficient conditions for full diversity, at the primary link and then in section V we propose an efficient method to design the alignment set. Section VI gives a numerical evaluation of the performance of the proposed methods. Finally, section VII provides some conclusions.

Notations: Boldface capital letters denote matrices, boldface lowercase letters denote column vectors. The operations $(.)^H$ and $tr(.)$ represent the Hermitian transpose and the trace of a matrix. $\mathcal{N}_L[\mathbf{A}]$ and $\mathcal{N}_R[\mathbf{A}]$ denote the left and right null-space of matrix $\mathbf{A}$, respectively. The notation $\mathbf{A} \sim \mathbf{B}$ denotes that matrix $\mathbf{A}$ and $\mathbf{B}$ are identically distributed. Let $\mathbf{A}$ be a matrix then $\mathbf{A}(i)$ denotes the column i of $\mathbf{A}$ and $\mathbf{A}(i,k)$ entry (i,k). $diag(\mathbf{A}_1, \ldots, \mathbf{A}_N)$ is a block diagonal matrix with entries $\mathbf{A}_n$, $n \in \{1, \ldots, N\}$. $\lambda_m[\mathbf{A}]$ denotes the mth largest singular value of matrix $\mathbf{A}$ and $|\mathbf{A}|$ the matrix $\mathbf{A}$ determinant. $\|\mathbf{a}\|$ represents the norm of vector $\mathbf{a}$. For two functions $f(x)$ and $g(x)$, $f(x) \doteq g(x)$ means $\lim_{x \to \infty} f(x) = \lim_{x \to \infty} g(x)$ [24]. The symbols $"\dot{\geq}"$ and $"\dot{\leq}"$ are similarly defined.

II. SYSTEM MODEL

We consider a set of N small-cells within the coverage area of a macro-cell, both sharing the same spectrum, as shown in Fig. 1. The N small-cell access points (AP) are connected through a high speed backhaul network (e.g. fiber) to a central unit (CU). In this work, we consider the uplink, i.e. the case where the UTs transmit information to the corresponding BS and APs (UT n sends information to AP n) and that both the macro and small-cells are operating in time division duplex (TDD) mode. We assume that the transmit power of the primary and secondary users is constrained to P_0 and P_n, respectively. In the following the macro-UT is denominated by UT_0 and the secondary UTs (APs) by UT_n (AP_n), $n \in \{1, \ldots, N\}$. This convention is also considered for all variables, i.e. the index zero refers to the primary terminals (BS and macro-UT) and an index higher than zero denotes secondary terminals.

A detailed block diagram of the system to be considered is presented in Fig. 2. At the primary, we assume that the BS and the UT have M and K antennas, respectively. Due to physical constraints at the transceivers it is likely that the number of antennas at the user terminal will be lower than the number of antennas at the BS. As such, we consider in
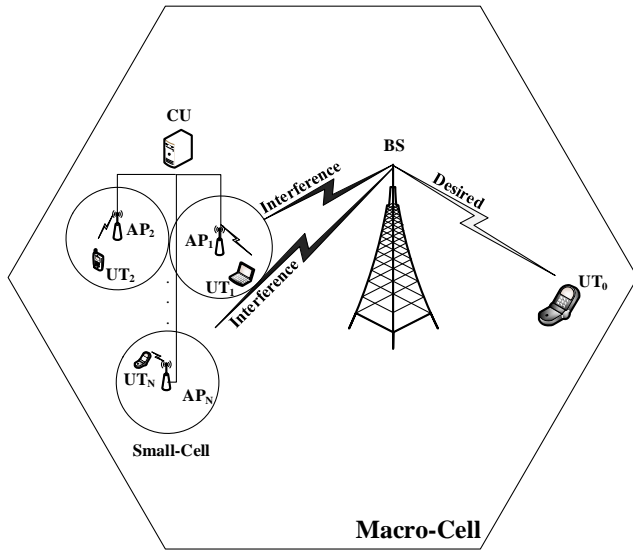
Fig. 1. Scenario: Set of N small-cells within the coverage area of a macro-cell.

the following that $K < M$. At the secondary, all terminals (both the UTs and APs) have M antennas [12]. The transmit signal ($\mathbf{x}_n$), at UT n, is obtained by a linear map ($\mathbf{V}_n$) of the vector of data symbols $\mathbf{d}_n$

$$\mathbf{x}_n = P_n^{1/2}\mathbf{V}_n\mathbf{d}_n \qquad (1)$$

The data symbols $\mathbf{d}_n$ are drawn from an uniform distribution over a Quadrature Amplitude Modulation (QAM) constellation with size B. The received signal at terminal n (BS is considered to be the receiver with index 0) can be expressed by

$$\mathbf{y}_n = \mathbf{H}_{n0}\mathbf{x}_0 + \sum_{i=1}^{N}\mathbf{H}_{ni}\mathbf{x}_i + \mathbf{s}_n \qquad (2)$$

where $\mathbf{x}_0$, $\mathbf{x}_i$, $\mathbf{s}_n$, $\mathbf{H}_{ni}$, denote the primary, the secondary UT$_i$ transmit signal, the zero mean white Gaussian noise with variance σ^2 and the channel between UT$_i$ and receiver n, respectively. We assume that the channel $\mathbf{H}_{ni}$ is complex Gaussian distributed with unitary variance. Channel $\mathbf{H}_{0n}$ is known at UT_n and channels $\mathbf{H}_{ni}, i \in \{0, \ldots, N\}$ are known at AP_n. Furthermore, channel $\mathbf{H}_{00}$ is known at the BS and macro-UT. These channels may be acquired by listening to the pilot signals broadcasted by the macro BS (macro UT) and small-cell UTs, since the macro cell and small cells mode of operation is TDD.

In the following we assume that the macro-cell link must remain free of inter-system interference. Furthermore, we consider that the macro UT uses all available channel dimensions ($min(M, K)$), since it is the licensee, i.e. it transmits K data streams. On the other hand, each secondary UT transmits only one data stream. To allow each secondary UT to transmit one data stream we consider that IA is used to align all transmissions in a common subspace. The required information for IA is made available at the secondary terminals either through the existence of a low-rate link between the two systems, or over the air through a dedicated
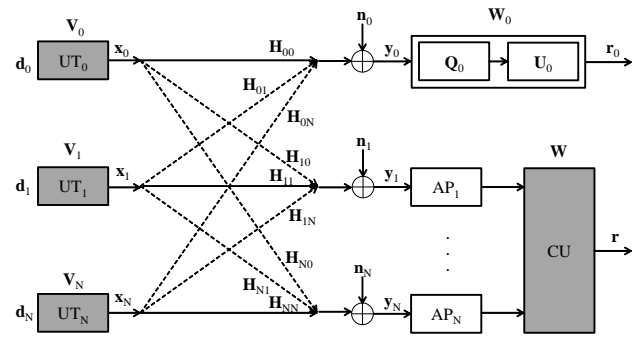

Fig. 2. Block diagram of the considered system.

cognitive pilot signal as proposed in [25] for the coexistence of various access technologies. It is important to minimize the resources required for this information, to avoid overloading the backhaul or the broadcast channel as they will also be used for other signaling information and not only this one.

A. Primary System

To decode the transmitted symbols the BS applies equalizer $\mathbf{W}_0 \in \mathbb{C}^{K \times M}$ to the received signal ($\mathbf{y}_0 \in \mathbb{C}^{M \times 1}$) to obtain $\mathbf{r}_0 \in \mathbb{C}^{K \times 1}$

$$\mathbf{r}_0 = \mathbf{W}_0\mathbf{y}_0 = \mathbf{W}_0\mathbf{H}_{00}\mathbf{x}_0 + \sum_{n=1}^{N}\mathbf{W}_0\mathbf{H}_{0n}\mathbf{x}_n + \mathbf{W}_0\mathbf{s}_0 \quad (3)$$

The equalizer is designed in two steps, see Fig. 2: first we apply the filtering matrix $\mathbf{Q}_0 \in \mathbb{C}^{(M-1) \times M}$ to remove the interference coming from the secondary network and then matrix $\mathbf{U}_0 \in \mathbb{C}^{K \times (M-1)}$ to remove the inter-stream interference. In the following we assume that $\mathbf{Q}_0\mathbf{Q}_0^H = \mathbf{I}$. The full equalizer is given by $\mathbf{W}_0 = \mathbf{U}_0\mathbf{Q}_0$. More details on the design of the filtering matrix $\mathbf{Q}_0$ are given in section III.

After secondary interference removal the received signal at the BS, is given by

$$\bar{\mathbf{r}}_0 = \mathbf{Q}_0\mathbf{y_0} = \mathbf{Q}_0\mathbf{H}_{00}\mathbf{x}_0 + \mathbf{Q}_0\mathbf{s}_0 \qquad (4)$$

Following the SVD analysis the equivalent channel matrix can be written as $\mathbf{H}_{00}^{eq} = \mathbf{Q}_0\mathbf{H}_{00} = \mathbf{U}_{00}^{eq}\mathbf{D}_{00}^{eq}\mathbf{V}_{00}^{eq\,H}$, where $\mathbf{D}_{00}^{eq} = diag(\lambda_1[\mathbf{Q}_0\mathbf{H}_{00}], \ldots, \lambda_K[\mathbf{Q}_0\mathbf{H}_{00}]) \in \mathbb{C}^{K \times K}$ is a diagonal matrix containing the singular values of $\mathbf{H}_{00}^{eq}$ and $\mathbf{U}_{00}^{eq} \in \mathbb{C}^{M \times K}$ ($\mathbf{V}_{00}^{eq} \in \mathbb{C}^{K \times K}$) denote the left (right) singular vectors of $\mathbf{H}_{00}^{eq}$. Assuming that the primary channel is perfectly known at both the BS and macro-UT we set $\mathbf{V}_0 = \mathbf{V}_{00}^{eq}$ and $\mathbf{U}_0 = \mathbf{U}_{00}^{eq\,H}$. This diagonalizes the equivalent channel and removes all inter-stream interference. According to (3) and the previously defined precoder/equalizer we have

$$\mathbf{r}_0 = P_0^{1/2}\mathbf{D}_{00}^{eq}\mathbf{d}_0 + \mathbf{n}_0 \qquad (5)$$

where $\mathbf{n}_0 = \mathbf{W}_0\mathbf{n}_0$ has the same statistical properties of $\mathbf{s}_0$ but with a reduced dimension, since $\mathbf{W}_0\mathbf{W}_0^H = \mathbf{U}_0\mathbf{Q}_0\mathbf{Q}_0^H\mathbf{U}_0^H = \mathbf{I}$. As a consequence the stream k SNR (SNR_k) is given by

$$\rho_k(\mathbf{H}_{00}, \mathbf{Q}_0, P_0) = \frac{P_0}{\sigma^2}\lambda_k[\mathbf{H}_{00}^H\mathbf{Q}_0^{\,H}\mathbf{Q}_0\mathbf{H}_{00}] \qquad (6)$$

B. Secondary System

At the secondary APs the received signal is forwarded to the CU for joint processing. As a consequence the CU has access to the signal $\mathbf{y}^H = [\mathbf{y}_1{}^H, \ldots, \mathbf{y}_N{}^H]^H (\in \mathbb{C}^{NM \times 1})$, to which it applies the linear transform $\mathbf{W} (\in \mathbb{C}^{(N+K) \times NM})$

$$\mathbf{r} = \mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{H}\mathbf{x} + \mathbf{W}\mathbf{s} \tag{7}$$

where $\mathbf{H} = [\mathbf{H}_1^H, \ldots, \mathbf{H}_n^H, \ldots, \mathbf{H}_N^H]^H (\in \mathbb{C}^{NM \times (K+NM)})$, $\mathbf{H}_n = [\mathbf{H}_{n0}, \ldots, \mathbf{H}_{ni}, \ldots, \mathbf{H}_{nN}] (\in \mathbb{C}^{M \times (K+NM)})$ denotes the concatenation of all APs channel matrices, $\mathbf{x} = [\mathbf{x}_0^H, \mathbf{x}_1^H, \ldots, \mathbf{x}_N^H]^H (\in \mathbb{C}^{(K+NM) \times 1})$ and $\mathbf{s} = [\mathbf{s}_1^H, \ldots, \mathbf{s}_N^H]^H (\in \mathbb{C}^{NM \times 1})$.

The CU joint processor has enough degrees of freedom (NM) to zero force the primary and secondary channels together, since only $N + K$ data streams are received at the CU. As a consequence we consider in the following that the CU applies a zero-forcing equalizer to the received signal, i.e. $\mathbf{W} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$, where $\mathbf{A} = \mathbf{H}\mathbf{V} (\in \mathbb{C}^{NM \times (N+K)})$ and $\mathbf{V} = diag(\mathbf{V}_0, \mathbf{V}_1, \ldots, \mathbf{V}_N) (\in \mathbb{C}^{(K+NM) \times (N+K)})$.

III. EQUALIZER AND ALIGNMENT SUBSPACE DESIGN

As the primary link must be free of inter-system interference the following condition must be respected

$$\mathbf{Q}_0 \mathbf{H}_{0n} \mathbf{V}_n = \mathbf{0}, \ n \in \{1, \ldots, N\} \tag{8}$$

Let us denote $\mathbf{v} = \alpha_n^{-1} \mathbf{H}_{0n} \mathbf{V}_n$ the alignment direction, for all $n \in \{1, \ldots, N\}$, where α_n denotes a positive constant.

Definition 1. *An alignment direction is a column vector ($\mathbf{v} \in \mathbb{C}^{M \times 1}$) that specifies completely the generated small-cell interference, from the point of view of the macro-receiver.*

As a consequence, the secondary precoders are set to

$$\begin{cases} \mathbf{V}_n = \alpha_n \mathbf{H}_{0n}^{-1} \mathbf{v} \\ \alpha_n = \left(\frac{P_n}{\mathbf{v}^H (\mathbf{H}_{0n} \mathbf{H}_{0n}^H)^{-1} \mathbf{v}} \right)^{\frac{1}{2}} \end{cases} \tag{9}$$

This value for constant α_n ensures that the power constraint P_n is respected at UT_n.

It should be mentioned that this approach may be extended for the case where the small cell UTs have a number of antennas less than M. In such a case the precoder must be replaced by $\mathbf{V}_n = \mathcal{N}_R[\mathbf{Q}_0 \mathbf{H}_{0n}]$. This leads to multiple values for the vectors $\mathbf{H}_{0n}\mathbf{V}_n, (n = 1, \ldots, N)$, although they all lie in the null space defined by (8). Simulation results indicated the same type of behavior for both cases but the asymptotic analysis, of section IV, addresses the specific case where the UTs have the same number of antennas as the BS.

Equation (8) can be rewritten as follows

$$\mathbf{Q}_0 \mathbf{v} = \mathbf{0} \tag{10}$$

From (10) we verify that the alignment direction is equal to the right-null space of $\mathbf{Q}_0$

$$\mathbf{v} = \mathcal{N}_R[\mathbf{Q}_0] \tag{11}$$

Definition 2. *The SVD of matrix $\mathbf{A} \in \mathbb{C}^{M \times K}$ is $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^H$, where $\mathbf{D} \in \mathbb{C}^{M \times K}$ and $\mathbf{U} \in \mathbb{C}^{M \times M}$ ($\mathbf{V} \in \mathbb{C}^{K \times K}$) denote a rectangular diagonal matrix containing the ordered*

singular values of $\mathbf{A}$ and a unitary matrix containing the left (right) singular vectors of $\mathbf{A}$, respectively.

For $M > K$ the right null-space of $\mathbf{A}$ is empty and the left null-space is made up of the last $M - K$ columns of $\mathbf{U}$ ($\mathcal{N}_L[\mathbf{A}] = [\mathbf{U}(K+1), \ldots, \mathbf{U}(M)]$).

For $M < K$ the left null-space of $\mathbf{A}$ is empty and the right null-space is made up of the last $K - M$ columns of $\mathbf{V}$ ($\mathcal{N}_R[\mathbf{A}] = [\mathbf{V}(M+1), \ldots, \mathbf{V}(K)]$).

From (9) and (11) we verify that the design of UT_n precoder $\mathbf{V}_n$ and alignment direction $\mathbf{v}$ are dependent on the design of the macro-BS filtering matrix $\mathbf{Q}_0$. To limit the amount of information exchange between the two systems we assume that the primary filtering matrix is selected from a finite set $\mathcal{T} = \{\mathbf{Q}_0^1, \ldots, \mathbf{Q}_0^t, \ldots, \mathbf{Q}_0^T\}$, with cardinality T, or equivalently that the alignment direction $\mathbf{v}$ is selected from set $\mathcal{A} = \{\mathbf{v}_1, \ldots, \mathbf{v}_T\}$. These two sets are related by (11). Furthermore, they are fixed and assumed to be known by the primary and secondary terminals.

For a given $\mathbf{Q}_0 \in \mathcal{T}$ and channel realization $\mathbf{H}_{00}$ the system bit-error-rate (BER), at the macro-BS, can be upper bounded, when using Gray encoding procedures as [26], [27]

$$BER(\mathbf{H}_{00}, \mathbf{Q}_0, P_0) \leq \frac{1}{K} \sum_{k=1}^{K} \mathcal{Q}\left(\sqrt{\xi \rho_k(\mathbf{H}_{00}, \mathbf{Q}_0, P_0)} \right) \tag{12}$$

where $\mathcal{Q}(x) = 1/\sqrt{2\pi} \int_x^{+\infty} e^{-t^2/2} dt$ is the Q-function and ξ depend on the constellation. For instance, when using a B-QAM modulation $\xi = 3/(B-1)$ [26], [27].

At each channel realization the macro-BS selects one element from set $\mathcal{T}$ as its filtering matrix. In the following, we (12), as a metric to select the optimum $\mathbf{Q}_0$. Therefore, by taking into account the filtering matrix selection process the system BER is

$$BER(\mathbf{H}_{00}, P_0, \mathcal{T}) = \min_{\mathbf{Q}_0 \in \mathcal{T}} BER(\mathbf{H}_{00}, \mathbf{Q}_0, P_0) \tag{13}$$

The average system BER is the expectation over the channel realizations of $BER(\mathbf{H}_{00}, P_0, \mathcal{T})$

$$\overline{BER}(P_0, \mathcal{T}) = \mathbb{E}_{\mathbf{H}_{00}} [BER(\mathbf{H}_{00}, P_0, \mathcal{T})] \tag{14}$$

As mentioned before, the alignment direction $\mathbf{v}$ (filter matrix $\mathbf{Q}_0$) is selected at the BS, using the primary BER (see (12)), as a metric. After, selecting the alignment direction $\mathbf{v}$ from set $\mathcal{A}$, the BS must send this selection to the secondary terminals so that they can use this information to form their precoders (see (9)). To do this, we consider that the BS sends the required information either through a low-rate cooperation link or a broadcast channel as previously mentioned.

IV. SUFFICIENT CONDITIONS FOR FULL DIVERSITY

In this section we present sufficient conditions on the alignment direction set to achieve full diversity order, at the primary link. Namely, we study the impact of the cardinality of set $\mathcal{T}$ on the BER performance of the primary link.

The use of multiple antennas at the transmitter and/or receiver provides multiple paths to the information signal to pass through. This provides multiple independently faded

replicas of the data symbol, at the receiver end, hence more reliable reception is achieved [24]. Intuitively the diversity order is equal to the number of independent channel paths the information signal passes through [24]. To formalize it we have the following definition [24]

Definition 3. *The primary system achieves a diversity c if its BER follows*

$$\lim_{P_0 \to \infty} -\frac{log\left(\overline{BER}(P_0, \mathcal{T})\right)}{log(P_0)} = c \quad (15)$$

It is well known that for a $M \times K$ MIMO system, with K data streams, the diversity order is equal to $M - K + 1$ [28]. In the following we say that the primary system achieves full diversity when the attained diversity is the same as for a system without interfering secondary network. Next we provide conditions for set $\mathcal{T}$ so that full-diversity is achieved.

Before going into the details for the case where the filter matrix is chosen from a finite set let us consider the case where $\mathbf{Q}_0$ can take any value from the complex field, i.e. when $T \to \infty$, which represents an upper bound for the finite set case. We call this scenario *Full-Info* in the following.

Definition 4. *The SVD of channel $\mathbf{H}_{00} \in \mathbb{C}^{M \times K}$ is $\mathbf{H}_{00} = \mathbf{U}_{00}\mathbf{D}_{00}\mathbf{V}_{00}^H$, where $\mathbf{U}_{00} = [\mathbf{U}_{00}^1, \ \mathbf{U}_{00}^0]$ is a unitary matrix containing the left singular vectors of the channel, $\mathbf{D}_{00} = [\mathbf{D}_{00}^{1\ H} \mathbf{0}]^H$ is a rectangular diagonal matrix containing the singular values of $\mathbf{H}_{00}$, $\mathbf{U}_{00}^0 = \mathcal{N}_L[\mathbf{H}_{00}] \in \mathbb{C}^{M \times (M-K)}$ is the left null-space of $\mathbf{H}_{00}$ and $\mathbf{U}_{00}^1 \in \mathbb{C}^{M \times K}$ the left singular vectors associated to the K nonzero singular values contained in diagonal matrix $\mathbf{D}_{00}^1 \in \mathbb{C}^{K \times K}$.*

Theorem 1. *The maximum diversity order is equal to $M - K + 1$ for a coexistence scenario, where a set of N small cells $(M \times M)$, each transmitting a stream of information, share the same spectrum with a macro-cell $(M \times K)$ delivering K streams of information. For the Full-Info scenario and $N > 0$ the SNR of stream $k \in \{1, \ldots, K\}$ is the same as the SNR of stream k for $N = 0$ (only primary transmits).*

Proof. See Appendix A. □

Theorem 1 defines an upper bound on the maximum diversity that can be achieved with limited exchange of information. Indeed, it shows that full-diversity can be achieved for the coexistence scenario, when the primary and secondary are able to exchange any amount of information they desire. Are we still able to achieve full-diversity $(M-K+1)$ when T is finite? In the following we analyse the limited information exchange case, but before providing the main result (Theorem 2), we present three lemmas that will be used in its proof.

Definition 5. *Let matrices $\mathbf{H}_{00} \in \mathbb{C}^{M \times K}$ and $\mathbf{G}_{00} \in \mathbb{C}^{M \times K}$ define two channels. We write $\mathbf{H}_{00} \succeq \mathbf{G}_{00}$ if both channels have the same left and right singular vectors but $\lambda_k[\mathbf{H}_{00}] \geq \lambda_k[\mathbf{G}_{00}]$ for all $k \in \{1, 2, \ldots, K\}$.*

Lemma 1. *Let $\mathbf{Q}_0 \in \mathbb{C}^{(M-1) \times M}$ be an arbitrary matrix. If $\mathbf{H}_{00} \succeq \mathbf{G}_{00}$ and $\lambda_k[\mathbf{G}_{00}] = \lambda_K[\mathbf{H}_{00}]$ for*

all $k \in \{1, 2, \ldots, K\}$, then $BER(\mathbf{H}_{00}, \mathbf{Q}_0, P_0) \leq BER(\mathbf{G}_{00}, \mathbf{Q}_0, P_0)$.

Proof. See Appendix B □

Lemma 2. *Consider a filtering matrix $\mathbf{Q}_0 \in \mathbb{C}^{(M-1) \times M}$, with the property $\mathbf{Q}_0\mathbf{Q}_0^H = \mathbf{I}$ two channels ($\mathbf{H}_{00}$ and $\mathbf{G}_{00}$) with $\mathbf{H}_{00} \succeq \mathbf{G}_{00}$ and $\lambda_k[\mathbf{G}_{00}] = \lambda_K[\mathbf{H}_{00}]$, for all $k \in \{1, 2, \ldots, K\}$, and an alignment direction $\mathbf{v} = \mathcal{N}_R[\mathbf{Q}_0]$. Then, the application of filtering matrix $\mathbf{Q}_0$ to $\mathbf{G}_{00}$ only affects the SNR of one of its K streams*

$$\rho_k(\mathbf{G}_{00}, \mathbf{Q}_0, P_0)$$
$$= \begin{cases} \frac{P_0}{\sigma^2}\lambda_K^2[\mathbf{H}_{00}], \ \forall \ k \in \{1, 2, \ldots, K-1\} \\ \frac{P_0}{\sigma^2}\lambda_K^2[\mathbf{H}_{00}](\mathbf{v}^H\mathcal{N}_L[\mathbf{H}_{00}]\mathcal{N}_L[\mathbf{H}_{00}]^H\mathbf{v}), \ k = K \end{cases}$$
$$(16)$$

where $\mathcal{N}_L[\mathbf{H}_{00}]^H\mathbf{v}\mathbf{v}^H\mathcal{N}_L[\mathbf{H}_{00}]$ is positive and always lower or equal than 1.

Proof. See Appendix C □

Corollary 1. *Consider a filtering matrix $\mathbf{Q}_0 \in \mathbb{C}^{(M-1) \times M}$, with the property $\mathbf{Q}_0\mathbf{Q}_0^H = \mathbf{I}$, two channels ($\mathbf{H}_{00}$ and $\mathbf{G}_{00}$) with $\mathbf{H}_{00} \succeq \mathbf{G}_{00}$ and $\lambda_k[\mathbf{G}_{00}] = \lambda_K[\mathbf{H}_{00}]$, for all $k \in \{1, 2, \ldots, K\}$, and an alignment direction $\mathbf{v} = \mathcal{N}_R[\mathbf{Q}_0]$. Then, the best alignment direction within set $\mathcal{A}$ is*

$$\mathbf{v} = arg \max_{\mathbf{v} \in \mathcal{A}} \mathbf{v}^H\mathcal{N}_L[\mathbf{H}_{00}]\mathcal{N}_L[\mathbf{H}_{00}]^H\mathbf{v} \quad (17)$$

Corollary 1 presents a suboptimal method to select the best alignment direction of set $\mathcal{A}$. This selection process only needs as inputs the corresponding primary link channel null-space ($\mathcal{N}_L[\mathbf{H}_{00}]$) and alignment set $\mathcal{A}$. Equation (17) indicates that the best alignment direction is the one which is closer to the vector describing the null space of the considered channel.

Lemma 3. *Let $\mathbf{U}_{00} \in \mathbb{C}^{M \times M}$ be a uniformly distributed unitary matrix (Haar-distributed) [29] and $\mathbf{D}_i \in \mathbb{C}^{M \times M}$ ($i \in \{1, 2\}$) two diagonal matrices, with $\mathbf{D}_i(m, m) = \{0, 1\}, m \in \{1, \ldots, M\}$, $tr(\mathbf{D}_1) = M - K$ and $tr(\mathbf{D}_2) = 2$. Let $\mathbf{h}_1 = \mathbf{\Phi}_1/\|\mathbf{\Phi}_1\|$ and $\mathbf{h}_2 = \mathbf{\Phi}_2/\|\mathbf{\Phi}_2\|$, where $\mathbf{\Phi}_1 \in \mathbb{C}^{M \times 1}$ and $\mathbf{\Phi}_2 \in \mathbb{C}^{(M-1) \times 1}$ are i.i.d complex Gaussian with zero mean and unitary variance, then*

$$tr(\mathbf{D}_1\mathbf{U}_{00}^H\mathbf{D}_2\mathbf{U}_{00}) \geq 1 - (1 - x)(1 - y) \quad (18)$$

where $x = \sum_{m=1}^{M-K} |\mathbf{h}_1(m)|^2$ and $y = \sum_{m=1}^{M-K} |\mathbf{h}_2(m)|^2$ have the following joint distribution

$$f_{X,Y}(x, y)$$
$$\leq \frac{M-1}{K-1}\left(\frac{(M-2)!}{(M-K-1)!(K-2)!}\right)^2(1-y)(xy)^{M-K-1}$$
$$(19)$$

Proof. See Appendix D □

With the support of the previous three lemmas we now present sufficient conditions on the set of alignment directions ($\mathcal{A}$) to achieve full-diversity order.

Theorem 2. *Consider a coexistence scenario, where a set of N small cells $(M \times M)$ share the same spectrum with a*

macro-cell ($K \times M$). Denote by L the maximum number of linearly independent vectors in the alignment set $\mathcal{A}$.

If $L = 1$ then full diversity is not achievable. For this case the diversity order is $M - K$ and the BER behaves as

$$\overline{BER}(P_0, \mathcal{T}) \doteq \left(\frac{\sigma^2}{P_0} \right)^{M-K} \tag{20}$$

If $L \geq 2$ then full diversity order is obtained, at the macro-cell link. More specifically, for $K < M - 1$ or $L > 2$ the BER scales as

$$\overline{BER}(P_0, \mathcal{T}) \doteq \left(\frac{\sigma^2}{P_0} \right)^{M-K+1} \tag{21}$$

For the specific case of $K = M - 1$ and $L = 2$ the BER behaves as

$$\overline{BER}(P_0, \mathcal{T}) \doteq log\left(\frac{P_0}{\sigma^2} \right) \left(\frac{\sigma^2}{P_0} \right)^{M-K+1} \tag{22}$$

Proof. See Appendix E. □

According to theorem 2 we verify that for both cases, $L = 2$ and $L > 2$, the alignment set defined by $\mathcal{A}$ achieves full diversity. Nevertheless, for $L = 2$ and $K = M - 1$ there is a $log(P_0/\sigma^2)$ term that affects the actual slope of the BER curve for moderate SNRs. Therefore, an alignment set with description length of only 1 bit is enough to obtain the same diversity as an alignment set with infinite cardinality.

V. Equalizer Set Optimization

In this section we propose a method to design the alignment set $\mathcal{A}$, using as a performance metric the average system BER.

Definition 6. *Consider the alignment set $\mathcal{A} = \{\mathbf{v}_1, \ldots, \mathbf{v}_T\}$. Let $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_T]$ be the concatenation of all alignment directions, from set $\mathcal{A}$. The SVD of $\mathbf{V}$ is $\mathbf{U_V} \mathbf{\Lambda_V} \mathbf{W_V}^H$, where $\mathbf{U_V} \in \mathbb{C}^{M \times M}$, $\mathbf{V_V} \in \mathbb{C}^{T \times T}$, $\mathbf{\Lambda_V} = [\mathbf{\Lambda_V^1}\ \mathbf{0}] \in \mathbb{C}^{M \times T}$ and $\mathbf{\Lambda_V^1} \in \mathbb{C}^{M \times M}$.*

Lemma 4. *For $L \geq M$ the primary link BER is upper bounded by*

$$\overline{BER}(P_0, \mathcal{T}) \dot{\leq} |\mathbf{VV}^H|^{-\frac{M-K+1}{M}} \left(\frac{P_0}{\sigma^2} \right)^{-(M-K+1)} \tag{23}$$

Proof. See Appendix F □

To optimize the alignment set we consider the average BER upper bound given by lemma 4. As a consequence, the alignment set optimization problem is

$$\max_{\mathbf{V}} |\mathbf{VV}^H|\ \ s.t.\ \mathbf{V}(t)^H \mathbf{V}(t) = 1,\ t \in \{1, \ldots, T\} \tag{24}$$

The constraint $\mathbf{V}(t)^H \mathbf{V}(t)$ follows from (11). In the following we denominate by *DET* method the process used to obtain a solution to optimization problem (24). Due to the multiple quadratic constraints in optimization problem (24), it is quite difficult to obtain a closed form solution. Nevertheless, the following theorem simplifies significantly this problem and allows to obtain such a solution.

Theorem 3. *Consider the following optimization problem*

$$\max_{\mathbf{V}} |\mathbf{VV}^H|\ \ s.t. \sum_{t=1}^{T} \mathbf{V}(t)^H \mathbf{V}(t) = T \tag{25}$$

If $T \geq M$ then optimization problems (24) and (25) have the same optimum value $((T/M)^M)$ which is achieved for $\mathbf{\Lambda}_V = \left[\sqrt{T/M} \mathbf{I}_M\ \mathbf{0} \right]$. The left and right singular vectors $\mathbf{U_V}$ and $\mathbf{W_V}$ have no impact on the optimum value of both problems, but for problem (24) $\mathbf{W_V}$ must be set according to constraints $\mathbf{V}(t)^H \mathbf{V}(t) = 1$, $t \in \{1, \ldots, T\}$.

Proof. See Appendix G □

The task of generating matrix $\mathbf{W_V}$ may be carried out with the help of algorithm 3.1 from [30]. To obtain a solution for (24) we use the following procedure

1) Set $\mathbf{\Lambda}_V = \left[(T/M)^{\frac{1}{2}} \mathbf{I}_M,\ \mathbf{0} \right]$ and $\mathbf{D} = \mathbf{\Lambda}_V^H \mathbf{\Lambda}_V$
2) Use algorithm 3.1 from [30] to generate $\mathbf{V}^H \mathbf{V}$, with spectrum given by $\mathbf{D}$.
3) Compute the SVD of $\mathbf{V}^H \mathbf{V} = \mathbf{U}_{\mathbf{V}^H \mathbf{V}} \mathbf{D} \mathbf{U}_{\mathbf{V}^H \mathbf{V}}^H$
4) Set $\mathbf{V} = \mathbf{\Lambda}_V \mathbf{U}_{\mathbf{V}^H \mathbf{V}}^H$

A. Alignment Set Examples

The procedure presented in the previous section generates an alignment set which respects all optimal conditions of problem (24). Nevertheless, there are other methods that allow us to achieve the same results. Indeed, in the following we present two different approaches to generate alignment set matrices with the same properties as the previous procedure. These, due to its simplicity may be of more interest for practical applications.

1) Two-Bit channel quantization: Let us consider that $\mathbf{V} \in \mathbb{C}^{2^{2M} \times M}$ is given by $\mathbf{V} = \alpha[(1 - 2\mathbf{Z}_r) + i(1 - 2\mathbf{Z}_i)]$ and $\mathbf{Z} = [\mathbf{Z}_r, \mathbf{Z}_i]$. Row n of matrix $\mathbf{Z} \in \{0,1\}^{2^{2M} \times 2M}$ contains the binary representation of integer n. We can think of the described matrix $\mathbf{V}$ as the quantization of the real and imaginary parts of the null-space of channel matrix $\mathbf{H}_{00}$, using one bit for each component. Matrix $\mathbf{V}$ contains redundant entries since row $\mathbf{V}_i$ results in the same SNR as row $-\mathbf{V}_i$, $j\mathbf{V}_i$ and $-j\mathbf{V}_i$. Therefore only one fourth of matrix $\mathbf{V}$ is needed, i.e. rows $-\mathbf{V}_i$, $j\mathbf{V}_i$ and $-j\mathbf{V}_i$ may be removed. Let us denote by $\mathbf{C}$ the matrix $\mathbf{V}$ after redundant row removal. As a consequence $\mathbf{C}^H \mathbf{C} = \alpha^2 2^{2M-1} \mathbf{I}_M$ and $\mathbf{C}(i)\mathbf{C}(i)^H = 2\alpha^2 M$. If we set α to $(2M)^{-1/2}$ then matrix $\mathbf{C}$ respects all conditions to achieve the maximum of problem (24).

2) DFT matrix: Matrix $\mathbf{V}$ can also be constructed by extracting M columns from a DFT matrix of dimension T. For this case we can also show that the resulting matrix respects all the optimal conditions of optimization problem (24).

VI. Performance Results

In this section we first present Monte Carlo simulations to verify the sufficient conditions for alignment set $\mathcal{A}$, i.e. theorem 2. For this numerical analysis we consider that the alignment set is generated by selecting T columns from the 4×4 identity matrix. Then, we assess the performance of the proposed *DET* method and compare it with the

Fig. 3. Average BER at the primary system, for QPSK. $M = 4$, $K \in \{2, 3\}$, $N = 5$, $T \in \{1, 2, 3, 4\}$ and Full-Info Scenario.
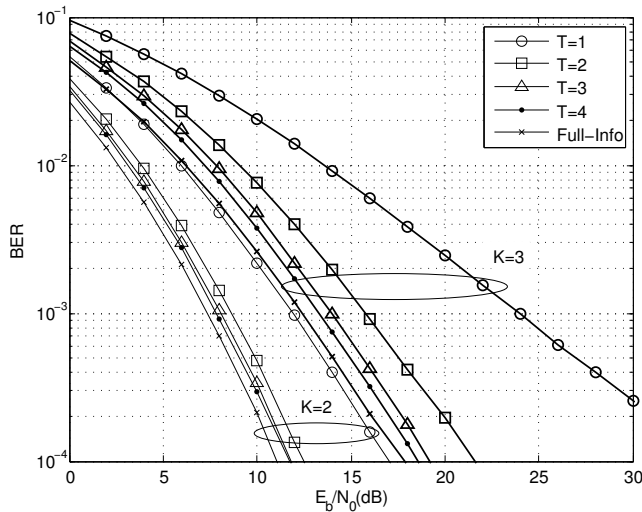

Fig. 4. Average BER at the primary system, for 16-QAM. $M = 4$, $K \in \{2, 3\}$, $N = 5$, $T \in \{1, 2, 3, 4\}$ and Full-Info Scenario.

RVQ approach [22], [31], where the alignment vectors are generated independently from a uniform distribution on the complex unit sphere and to the Grassmannian packings method (denominated Grass method in the following) with codebooks optimized using the chordal distance as a metric [20], [32]. In the simulations we have considered normalized channels, which means that the average long term gain is equal to one. However the derivations were made for arbitrary transmit powers at the primary and secondary (P_0 and P_n) and so the inclusion of the path loss aspect is equal to the change of the values of P_0 and P_n in the simulations. This will only shift the BER curves, since all interference is completely removed (the two systems may be considered independent after interference removal). In the following we consider identical noise power spectral density (N_o) values for both the primary and secondary systems. This assumption enables us to readily verify from the results the impact of the power imbalance between the primary and secondary systems.

We consider a scenario with $N = 5$ APs, $M = 4$, $K \in \{2, 3\}$ and $B \in \{4, 16\}$, where B denotes the QAM modulation size. The power of the primary UT is assumed to be 10 dB higher than the one of secondary UTs. The results presented are a function of the bit error rate over the energy per bit to noise power spectral density ratio per antenna (E_b/N_0). At the BS the best filtering matrix (alignment direction) is selected using the BER as a metric. We consider in the following that the set of alignment directions is full rank. Recalling that T denotes the cardinality of alignment set $\mathcal{A}$, then for $T \leq M$ and $T > M$ the matrix $\mathbf{V}$ has rank $L = M$ and $L = T$, respectively. All Grassmannian codebooks used, in this section, have a gap lower than 0.01 from the Rankin bound [32]. Worst case results are presented for all three methods. To obtain these results 10000 different alignment sets were generated.

Let us start with the numerical verification of theorem 2. In Fig. 3 and 4 we present the BER results, at the primary link, for QPSK and 16-QAM modulations, respectively. Results show
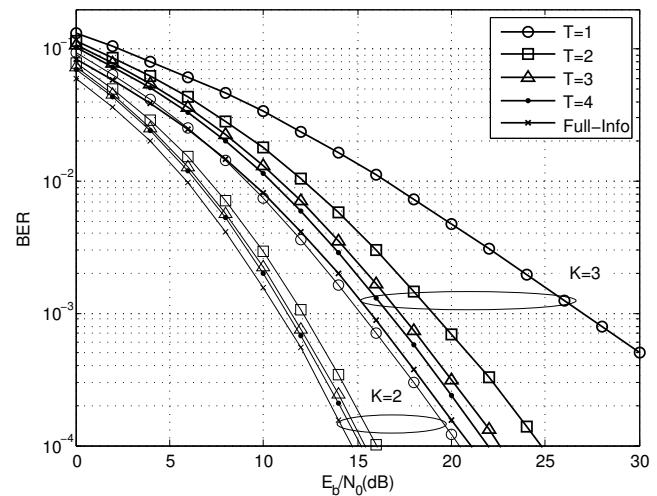
that for $T = 1$ the performance is degraded as the slope of the BER curve is the lowest. On the other hand, for $T \geq 2$ the slope improves significantly. Fig. 3 and 4 show that the slope of the BER curves for both $T = 3$ and $T = 4$ is the same, as expected from the demonstration in the previous section. Indeed, it is similar to the slope of the BER curve for the Full-Info scenario. Nevertheless, for $T = 2$ and $K = M - 1 = 3$ we can verify that the slope, for the SNR values considered, is lower than for $T > 2$ as demonstrated in section IV. The same is not true for $K = 2 < M - 1$, as shown in Theorem 2, where the slope is still the same as in the Full-Info scenario, i.e., $T = \infty$. This degeneracy, present for the case $T = 2$ and $K = M - 1$, stems from the logarithm term in (22), which slows down the convergence rate of the BER curve slope with the SNR. Comparing the results for QPSK (Fig. 3) and 16-QAM (Fig. 4) we verify that the behavior of the curves is similar (around a 4 dB shift between the two modulations, as expected). If we fix the codebook size (T), and increase the number of antennas at the macro UT (K), the gap to the Full-Info scenario increases. For $T = 4$ the gap is around 0.5 dB for $K = 2$ and increases to 1 dB for $K = 3$. This is due to the fact that at the BS, the interference is restricted to a subspace of an M-dimensional vector space with dimensionality 1 while the desired signal occupies a subspace of dimension K. As K increases the number of free space dimensions decreases and therefore the degrees of freedom available for selecting the interference subspace decrease.

Let us now consider the optimization of the alignment set $\mathcal{A}$ using the DET, RVQ and Grass methods. Fig. 5 and 6 present results for codebooks with size $T = 8$ optimized using the proposed scheme (DET method), the RVQ and the Grass methods. These results are only for QPSK modulation as the results for 16-QAM will be similar, around 4 dB worse. We have considered a codebook with moderate size, $T = 8$, since as the set cardinality increases more information is exchanged between both systems making the set generation method less important. Fig. 5 and 6 show the required E_b/N_0 to achieve a target BER of 10^{-3} versus $M \in \{4, 5, 6, 7, 8\}$ for $K = M - 2$ and $K = M - 1$, respectively. The selected values of K ensure a
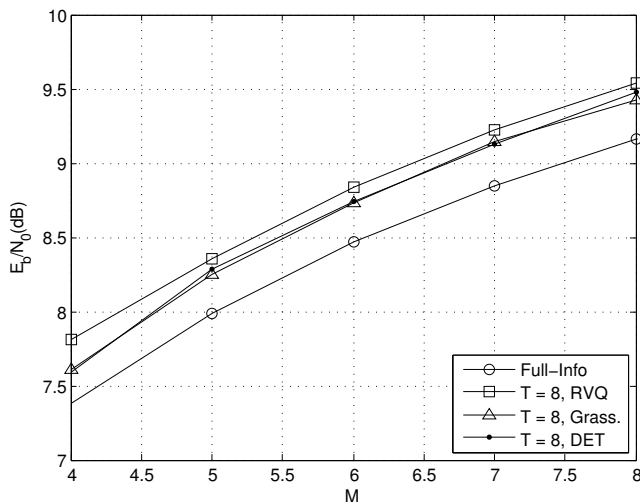
Fig. 5. E_b/N_0 needed to achieve a target BER of 10^{-3} for a codebook with size $T = 8$, $K = M - 2$ and $M \in \{4, 5, 6, 7, 8\}$.

Fig. 7. Average BER at the secondary system, for QPSK and 16-QAM, $M = 4$, $K \in \{2, 3\}$, $N = 5$.
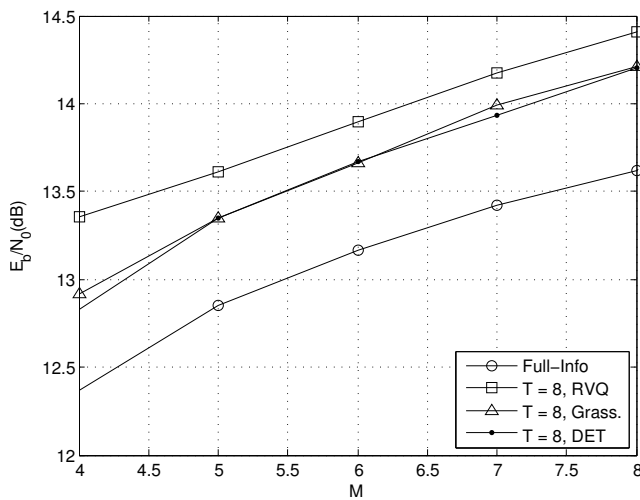

Fig. 6. E_b/N_0 needed to achieve a target BER of 10^{-3} for a codebook with size $T = 8$, $K = M - 1$ and $M \in \{4, 5, 6, 7, 8\}$.

fair comparison between the considered M values as for $K = M - 1$ ($K = M - 2$) the diversity order is kept fixed and equal to 2 (3). The performance of the proposed DET method is similar to the Grassmannian method, for both $K = M - 2$ and $K = M - 1$. In comparison to the RVQ method the proposed DET method achieves a gain of around 0.1 and 0.3 dB, for $K = M - 2$ and $K = M - 1$, respectively. On the other hand, the gap to the Full-Info scenario is around 0.3 and 0.5 dB.

Now, focusing on the secondary link, we see that in the high SNR the CU can remove completely the primary interference using the ZF equalizer. The performance for $K = 2$ is better than the one obtained for $K = 3$, since for the latter case more degrees of freedom are lost to remove the primary interference. As expected for 16-QAM the BER curve is shifted around 4 dB to the right.

VII. CONCLUSION

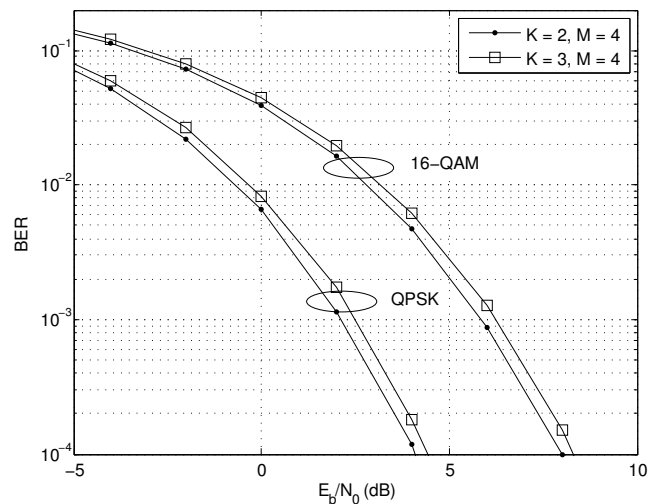In this manuscript we have considered the coexistence of a set of small-cells within a macro-cell. To limit the amount of information exchange between the two systems it was considered that the small-cells align their transmission along a common direction and that this direction is selected from a finite set. We have proposed an efficient method to construct the alignment set, for a given set cardinality, which has improved performance when compared to RVQ and similar performance to the Grassmanian packing method. Sufficient conditions for the alignment set where provided so that full diversity order is achieved, at the primary link. It was verified that an alignment set with just two alignment vectors is enough to achieve the full system diversity. This allows the primary and secondary systems to coexist with very limited information exchange between both systems. Indeed, with just 1 bit of information exchange the diversity of the primary is the same as in the case were an unlimited amount of information is exchanged. As a final conclusion we may say that the diversity of the primary network must match the diversity of the secondary network, i.e. the secondary network must increase the set of allowed transmit directions to improve the diversity of the primary system.

APPENDIX A
PROOF THEOREM 1

If $N = 0$ there are no small-cells only the macro-cell, then the considered scenario is just a single-user $M \times K$ MIMO communication system. For a single-user $M \times K$ MIMO system, using the SVD to decompose the channel in a set of K parallel channels the achieved diversity order is equal to $M - K + 1$ [28]. The SNR for stream k for this scenario is

$$\rho_k(\mathbf{H}_{00}, \mathbf{I}, P_0) = \frac{P_0}{\sigma^2} \lambda_k[\mathbf{H}_{00}] \tag{26}$$

Now let us consider a more general case, where $N > 0$. Consider that the small-cell terminals are using interference alignment to align all the generated interference along a common alignment direction $\mathbf{v}$. Assume the SVD of the primary channel as in definition 4 and set $\mathbf{v}$ equal to any column of $\mathbf{U}_{00}^0$ and the macro-cell receiver interference filtering matrix

$\mathbf{Q}_0 = \mathbf{U}_{00}^{1\,H}$. From the properties of $\mathbf{U}_{00}$ we verify that after the projection over $\mathbf{Q}_0$ the resulting equivalent primary channel is given by $\mathbf{D}_{00}^1 \mathbf{V}_{00}^H$. Therefore, the singular values of the original channel were not affected by the interference filtering process and the stream k SNR is

$$\rho_k(\mathbf{H}_{00}, \mathbf{U}_{00}^{1\,H}, P_0) = \frac{P_0}{\sigma^2} \lambda_k[\mathbf{H}_{00}] \tag{27}$$

the same as in the non-coexistence scenario. As a result, the diversity order for $N > 0$ is the same as for $N = 0$, i.e $M - K + 1$.

APPENDIX B
PROOF OF LEMMA 1

Let $\mathbf{Q}_0 \in \mathbb{C}^{(M-1) \times M}$ be an arbitrary matrix and $\mathbf{H}_{00} \succeq \mathbf{G}_{00}$ with $\lambda_k[\mathbf{G}_{00}] = \lambda_K[\mathbf{H}_{00}]$ for all $k \in \{1, 2, \ldots, K\}$ then as we have by construction $\mathbf{G}_{00}\mathbf{G}_{00}^H \preceq \mathbf{H}_{00}\mathbf{H}_{00}^H$ it follows from the properties of positive definite matrices [33] that $\lambda_m[\mathbf{Q}_0\mathbf{G}_{00}\mathbf{G}_{00}^H\mathbf{Q}_0^H] \leq \lambda_m[\mathbf{Q}_0\mathbf{H}_{00}\mathbf{H}_{00}^H\mathbf{Q}_0^H]$, for all $m \in \{1, \ldots, K\}$. Therefore the BER, see (12), for $\mathbf{H}_{00}$ is upper bounded by the BER of the link with channel matrix $\mathbf{G}_{00}$.

APPENDIX C
PROOF OF LEMMA 2

By assumption the SVD of channel $\mathbf{G}_{00}$ is

$$\begin{aligned}
\mathbf{G}_{00} &= \lambda_K[\mathbf{H}_{00}][\mathbf{U}_{00}^1 \; \mathbf{U}_{00}^0][\mathbf{I} \; \mathbf{0}]^H \mathbf{V}_{00}^H \\
&= \lambda_K[\mathbf{H}_{00}]\mathbf{U}_{00}^1\mathbf{V}_{00}^H
\end{aligned} \tag{28}$$

where $\mathbf{U}_{00}^0 = \mathcal{N}_L[\mathbf{H}_{00}]$. As a consequence we have the following

$$\mathbf{G}_{00}^H\mathbf{G}_{00} = \lambda_K^2[\mathbf{H}_{00}]\mathbf{I} \tag{29}$$

$$\mathbf{G}_{00}\mathbf{G}_{00}^H = \lambda_K^2[\mathbf{H}_{00}]\mathbf{U}_{00}^1\mathbf{U}_{00}^{1\,H} \tag{30}$$

As the matrix of left singular vectors is unitary it follows that $\mathbf{U}_{00}^1\mathbf{U}_{00}^{1\,H} + \mathbf{U}_{00}^0\mathbf{U}_{00}^{0\,H} = \mathbf{I}$, then

$$\mathbf{G}_{00}\mathbf{G}_{00}^H = \lambda_K^2[\mathbf{H}_{00}](\mathbf{I} - \mathbf{U}_{00}^0\mathbf{U}_{00}^{0\,H}) \tag{31}$$

The SVD of filter matrix $\mathbf{Q}_0 \in \mathbb{C}^{(M-1) \times M}$ is

$$\mathbf{Q}_0 = \mathbf{U}_{\mathbf{Q}_0}[\mathbf{D}_{\mathbf{Q}_0} \; \mathbf{0}][\mathbf{V}_{\mathbf{Q}_0}^1 \; \mathbf{v}_{\mathbf{Q}_0}^0]^H \tag{32}$$

where $\mathbf{v}_{\mathbf{Q}_0}^0 = \mathcal{N}_R[\mathbf{Q}_0] = \mathbf{v}$, see (11). As $\mathbf{Q}_0\mathbf{Q}_0^H = \mathbf{U}_{\mathbf{Q}_0}\mathbf{D}_{\mathbf{Q}_0}\mathbf{D}_{\mathbf{Q}_0}^H\mathbf{U}_{\mathbf{Q}_0}^H$, from (32) and $\mathbf{Q}_0\mathbf{Q}_0^H = \mathbf{I}$, by assumption, then $\mathbf{D}_{\mathbf{Q}_0}\mathbf{D}_{\mathbf{Q}_0}^H = \mathbf{I}$. Therefore, as $\mathbf{D}_{\mathbf{Q}_0}$ is diagonal then $\mathbf{D}_{\mathbf{Q}_0}^H\mathbf{D}_{\mathbf{Q}_0} = \mathbf{I}$, from (32) we have the following

$$\mathbf{Q}_0^H\mathbf{Q}_0 = \mathbf{V}_{\mathbf{Q}_0}^1\mathbf{V}_{\mathbf{Q}_0}^{1\,H} = \mathbf{I} - \mathbf{v}\mathbf{v}^H \tag{33}$$

The second equality follows from the fact that the right singular vectors of $\mathbf{Q}_0$ are unitary. From (29) and (33) it follows that

$$\begin{aligned}
\mathbf{G}_{00}^H\mathbf{Q}_0^H\mathbf{Q}_0\mathbf{G}_{00} &= \mathbf{G}_{00}^H(\mathbf{I} - \mathbf{v}\mathbf{v}^H)\mathbf{G}_{00} \\
&= \lambda_K^2[\mathbf{H}_{00}]\mathbf{I} - \mathbf{G}_{00}^H\mathbf{v}\mathbf{v}^H\mathbf{G}_{00}
\end{aligned} \tag{34}$$

As $\mathbf{G}_{00}^H\mathbf{v}\mathbf{v}^H\mathbf{G}_{00}$ is a rank 1 matrix, it has just one non-zero singular vector, i.e. its SVD is

$$\mathbf{G}_{00}^H\mathbf{v}\mathbf{v}^H\mathbf{G}_{00} = \mathbf{L}\Omega\mathbf{L}^H \tag{35}$$

where $\mathbf{L} \in \mathbb{C}^{K \times K}$ denote the corresponding singular vectors matrix, $\Omega = diag(\Omega_1, 0, \ldots, 0) \in \mathbb{C}^{K \times K}$ and $\Omega_1 = tr(\mathbf{G}_{00}^H\mathbf{v}\mathbf{v}^H\mathbf{G}_{00}) = \mathbf{v}^H\mathbf{G}_{00}\mathbf{G}_{00}^H\mathbf{v}$. Using (31) we obtain

$$\Omega_1 = \mathbf{v}^H\mathbf{G}_{00}\mathbf{G}_{00}^H\mathbf{v} = \lambda_K^2[\mathbf{H}_{00}](1 - \mathbf{v}^H\mathbf{U}_{00}^0\mathbf{U}_{00}^{0\,H}\mathbf{v}) \tag{36}$$

From (34) and (35) we get that the SVD of $\mathbf{G}_{00}^H\mathbf{Q}_0^H\mathbf{Q}_0\mathbf{G}_{00}$ is

$$\mathbf{G}_{00}^H\mathbf{Q}_0^H\mathbf{Q}_0\mathbf{G}_{00} = \mathbf{L}\Gamma\mathbf{L}^H \tag{37}$$

where $\Gamma = \lambda_K^2[\mathbf{H}_{00}]diag(1, 1, \ldots, \mathbf{v}^H\mathcal{N}_L[\mathbf{H}_{00}]\mathcal{N}_L[\mathbf{H}_{00}]^H\mathbf{v})$. As $\mathbf{v}$ is unitary and $\mathbf{U}_{00}^0$ is equal to the last $M - K$ columns of a unitary matrix, then $\mathbf{v}^H\mathcal{N}_L[\mathbf{H}_{00}]\mathcal{N}_L[\mathbf{H}_{00}]^H\mathbf{v} \leq 1$ and positive. As a consequence, from (6) we verify that the SNR of stream k, after the application of filtering matrix $\mathbf{Q}_0$, is given by (16).

APPENDIX D
PROOF LEMMA 3

Let $\mathbf{U}_{00} = [\mathbf{a}_1^H, \ldots, \mathbf{a}_m^H, \ldots, \mathbf{a}_M^H] \in \mathbb{C}^{M \times M}$, be a unitary matrix drawn uniformly from the ensemble of unitary matrices [29]. To generate a uniform random matrix $\mathbf{U}_{00}$ over the set of unitary matrices we may use the Gram-Schmidt process [29]. Using the Gram-Schmidt procedure vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are as follows

$$\mathbf{a}_1 = \frac{\mathbf{b}_1}{\|\mathbf{b}_1\|}, \quad \mathbf{b}_1 = \mathbf{g}_1 \tag{38}$$

$$\mathbf{a}_2 = \frac{\mathbf{b}_2}{\|\mathbf{b}_2\|}, \quad \mathbf{b}_2 = \mathbf{g}_2\mathbf{R}, \quad \mathbf{R} = \mathbf{I} - \mathbf{a}_1^H\mathbf{a}_1 \tag{39}$$

where $\mathbf{g}_1 \in \mathbb{C}^{1 \times M}$ and $\mathbf{g}_2 \in \mathbb{C}^{1 \times M}$ are i.i.d complex Gaussian random vectors with zero mean and unitary variance. Without loss of generality, assume that $\mathbf{D}_1 = diag(1, \ldots, 1, 0, \ldots, 0)$, $tr(\mathbf{D}_1) = M - K$ and $\mathbf{D}_2 = diag(1, 1, 0, \ldots, 0)$, then

$$\begin{aligned}
tr(\mathbf{D}_1\mathbf{U}_{00}^H\mathbf{D}_2\mathbf{U}_{00}) &= tr(\mathbf{D}_1(\mathbf{a}_1^H\mathbf{a}_1 + \mathbf{a}_2^H\mathbf{a}_2)) \\
&= \mathbf{a}_1\mathbf{D}_1\mathbf{a}_1^H + \mathbf{a}_2\mathbf{D}_1\mathbf{a}_2^H
\end{aligned} \tag{40}$$

Let $\mathbf{D_R} = diag(0, 1, \ldots, 1)$ and $\mathbf{U_R} = [\mathbf{a}_1^H, \mathcal{N}_R[\mathbf{a}_1]]$. $\mathbf{R}$ may be decomposed as $\mathbf{U_R}\mathbf{D_R}\mathbf{U_R}^H$. From the definition of the null-space operator $(\mathcal{N}_R[.])$ we have that $\mathbf{U_R}$ is unitary. It follows then that $\mathbf{b}_2 = \mathbf{g}_2\mathbf{R} \sim \mathbf{g}_2\mathbf{D_R}\mathbf{U_R}^H \sim \Phi_2\mathcal{N}_R[\mathbf{a}_1]^H$, where $\Phi_2 \in \mathbb{C}^{1 \times (M-1)}$ is a i.i.d random complex Gaussian vector with zero mean and unitary variance. Again, from the definition of the null-space operator $(\mathcal{N}_R[.])$ we have that $\mathcal{N}_R[\mathbf{a}_1]^H\mathcal{N}_R[\mathbf{a}_1] = \mathbf{I}$ then

$$\mathbf{a}_2\mathbf{D}_1\mathbf{a}_2^H \sim \frac{\Phi_2\mathbf{B}\Phi_2^H}{\Phi_2\Phi_2^H} \sim \frac{\Phi_2\mathbf{D_B}\Phi_2^H}{\Phi_2\Phi_2^H} \tag{41}$$

where $\mathbf{B} = \mathcal{N}_R[\mathbf{a}_1]^H\mathbf{D}_1\mathcal{N}_R[\mathbf{a}_1]$. The second equality follows from the SVD of $\mathbf{B} = \mathbf{U_B}\mathbf{D_B}\mathbf{U_B}^H$. Let $\mathbf{C} = \mathbf{D}_1^H\mathcal{N}_R[\mathbf{a}_1]\mathcal{N}_R[\mathbf{a}_1]^H\mathbf{D}_1$, them from the definition of $\mathbf{U_R}$

$$\begin{aligned}
\mathbf{C} &= \mathbf{D}_1^H(\mathbf{I} - \mathbf{a}_1^H\mathbf{a}_1)\mathbf{D}_1 = \mathbf{D}_1^H\mathbf{D}_1 - \mathbf{D}_1^H\mathbf{a}_1^H\mathbf{a}_1\mathbf{D}_1 \\
&= diag(\mathbf{I} - \mathbf{e}_1^H\mathbf{e}_1, \mathbf{0})
\end{aligned} \tag{42}$$

where $\mathbf{e}_1 = [\mathbf{a}_1(1), \ldots, \mathbf{a}_1(M-K)]$. As $\mathbf{e}_1^H\mathbf{e}_1$ is a rank 1 matrix, the nonzero singular values of $\mathbf{C}$ are equal to $\{1, \ldots, 1, 1 - \mathbf{e}_1\mathbf{e}_1^H\}$, where, by definition of $\mathbf{e}_1$, $\mathbf{e}_1\mathbf{e}_1^H = \mathbf{a}_1\mathbf{D}_1\mathbf{a}_1^H$. The nonzero singular values of $\mathbf{B}$ are equal to

the nonzero singular values of $\mathbf{C}$. Let $\mathbf{h}_1 \sim \mathbf{a}_1$ and $\mathbf{h}_2 = \mathbf{\Phi}_2/\|\mathbf{\Phi}_2\|$, then from (38) and (41) we have

$$
\begin{aligned}
tr(\mathbf{D}_1 \mathbf{U}_{00}^H \mathbf{D}_2 \mathbf{U}_{00}) &\sim x + (y - |\mathbf{h}_2(M-K)|^2) \\
&\quad + (1-x)|\mathbf{h}_2(M-K)|^2 \\
&\geq 1 - (1-x)(1-y)
\end{aligned}
\tag{43}
$$

where $x = \sum_{m=1}^{M-K} |\mathbf{h}_1(m)|^2$ and $y = \sum_{m=1}^{M-K} |\mathbf{h}_2(m)|^2$. Random variables x and y are independent. Using the definition of $\mathbf{h}_1$ and $\mathbf{h}_2$ and random variable transformation rules, the joint distribution of x and y ($f_{X,Y}(x,y)$) is upper bounded as in (19).

APPENDIX E
PROOF THEOREM 2

First we consider the case where $T = 1$. For this case we can assume without loss of generality that the alignment direction is equal to $\mathbf{v} = [1, 0, \ldots, 0]^H$. Therefore, to remove the secondary interference the BS must consider just the received signal at antennas 2 to M as the received signal at antenna 1 is not interference free. This implies that after the application of the filtering matrix the resulting channel is equivalent to that of a $(M-1) \times K$ MIMO system. As such, we are only able to achieve a diversity of $M - K$ [28], not $M - K + 1$ (full-diversity).

Now let us study the case where $T > 1$. Consider a channel $\mathbf{G}_{00}$ with $\mathbf{H}_{00} \succeq \mathbf{G}_{00}$ and $\lambda_k[\mathbf{G}_{00}] = \lambda_K[\mathbf{H}_{00}]$ for all $k \in \{1, 2, \ldots, K\}$. From Lemma 1 and for any filtering matrix $\mathbf{Q}_0$, $BER(\mathbf{H}_{00}, \mathbf{Q}_0, P_0) \leq BER(\mathbf{G}_{00}, \mathbf{Q}_0, P_0)$. Therefore, from (13) we get $BER(\mathbf{H}_{00}, P_0, \mathcal{T}) \leq BER(\mathbf{G}_{00}, P_0, \mathcal{T})$. By Lemma 2 the filtering process only affects the SNR of one of its K streams. From the same Lemma the minimum stream SNR is equal to

$$
\rho_K(\mathbf{G}_{00}, \mathbf{Q}_0, P_0) = \frac{P_0}{\sigma^2} \lambda_K^2[\mathbf{H}_{00}](\mathbf{v}^H \mathcal{N}_L[\mathbf{H}_{00}] \mathcal{N}_L[\mathbf{H}_{00}]^H \mathbf{v})
\tag{44}
$$

As all other $K-1$ streams have a better SNR than stream K the BER for channel $\mathbf{G}_{00}$ is upper bounded by, see (13)

$$
\begin{aligned}
BER(\mathbf{G}_{00}, P_0, \mathcal{T}) &\leq \min_{\mathbf{v} \in \mathcal{A}} \mathcal{Q}\left(\sqrt{\xi \rho_K(\mathbf{G}_{00}, \mathbf{Q}_0, P_0)}\right) \\
&= \mathcal{Q}\left(\sqrt{\xi \mu \frac{P_0}{\sigma^2} \lambda_K^2[\mathbf{H}_{00}]}\right)
\end{aligned}
\tag{45}
$$

where $\mu = \max_{\mathbf{v} \in \mathcal{A}} \{\mathbf{v}^H \mathcal{N}_L[\mathbf{H}_{00}] \mathcal{N}_L[\mathbf{H}_{00}]^H \mathbf{v}\}$ represents the loss incurred by the fact that the alignment direction $\mathbf{v}$ is selected from a finite set $\mathcal{A}$ with cardinality T. The equality follows from the monotonic properties of the Q-function.

By Theorem II.2 from [34], $\mathbf{H}_{00}$ has its left and right singular vectors statistically independent of its singular values. Hence, the average BER of channel $\mathbf{G}_{00}$ is

$$
\overline{BER}(P_0, \mathcal{T}) \leq \mathbb{E}_{\lambda_K[\mathbf{H}_{00}], \mu}\left[\mathcal{Q}\left(\sqrt{\xi \mu \frac{P_0}{\sigma^2} \lambda_K^2[\mathbf{H}_{00}]}\right)\right]
\tag{46}
$$

As the maximum value of a given sequence is always higher than its average value

$$
\begin{aligned}
\mu &= \max_{\mathbf{v} \in \mathcal{A}} \{\mathbf{v}^H \mathcal{N}_L[\mathbf{H}_{00}] \mathcal{N}_L[\mathbf{H}_{00}]^H \mathbf{v}\} \\
&\geq \frac{1}{T} tr(\mathcal{N}_L[\mathbf{H}_{00}]^H \mathbf{V} \mathbf{V}^H \mathcal{N}_L[\mathbf{H}_{00}])
\end{aligned}
\tag{47}
$$

where $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_T]$ is the concatenation of all alignment directions from the alignment set $\mathcal{A}$. Consider the SVD of $\mathbf{V}\mathbf{V}^H = \mathbf{U}_{\mathbf{V}} \mathbf{D}_{\mathbf{V}} \mathbf{U}_{\mathbf{V}}^H$, where $\mathbf{D}_{\mathbf{V}}$, $\mathbf{U}_{\mathbf{V}} \in \mathbb{C}^{M \times M}$. $\mathcal{N}_L[\mathbf{H}_{00}]$ is equal to the last $M - K$ columns of the unitary matrix $\mathbf{U}_{00}$. As $\mathbf{U}_{00} \in \mathbb{C}^{M \times M}$ is uniformly distributed over the set of unitary matrices [34] and $\mathbf{U}_{\mathbf{V}}$ is unitary then $\mathbf{U}_{\mathbf{V}}^H \mathcal{N}_L[\mathbf{H}_{00}]$ has the same distribution as $\mathcal{N}_L[\mathbf{H}_{00}]$. Consequently, only the eigenvalues of $\mathbf{V}\mathbf{V}^H$ matter in upper bound (47)

$$
\begin{aligned}
\mu &\leq \frac{1}{T} tr(\mathcal{N}_L[\mathbf{H}_{00}]^H \mathbf{V} \mathbf{V}^H \mathcal{N}_L[\mathbf{H}_{00}]) \\
&\sim \frac{1}{T} tr(\mathcal{N}_L[\mathbf{H}_{00}]^H \mathbf{D}_{\mathbf{V}} \mathcal{N}_L[\mathbf{H}_{00}])
\end{aligned}
\tag{48}
$$

Let $\kappa = \frac{1}{T} tr(\mathcal{N}_L[\mathbf{H}_{00}]^H \mathbf{D}_{\mathbf{V}} \mathcal{N}_L[\mathbf{H}_{00}])$ then from (46) and (48) the BER is upper bounded by

$$
\overline{BER}(P_0, \mathcal{T}) \leq \mathbb{E}_{\lambda_K[\mathbf{H}_{00}], \kappa}\left[\mathcal{Q}\left(\sqrt{\xi \kappa \frac{P_0}{\sigma^2} \lambda_K^2[\mathbf{H}_{00}]}\right)\right]
\tag{49}
$$

Using the exponential upper bound of the Q-function ($Q(x) \leq 1/2 e^{-x^2/2}$) and evaluating the expectation operator of (49) over $\lambda_K[\mathbf{H}_{00}]$ we get (see [35] equations (127) and (128))

$$
\begin{aligned}
\overline{BER}(P_0, \mathcal{T}) &\dot{\leq} \mathbb{E}_{\kappa}\left[\left(\kappa \frac{P_0}{\sigma^2}\right)^{-(M-K+1)}\right] \\
&= \mathbb{E}_{\kappa}\left[\kappa^{-(M-K+1)}\right]\left(\frac{P_0}{\sigma^2}\right)^{-(M-K+1)}
\end{aligned}
\tag{50}
$$

As a consequence if the expectation $\mathbb{E}_{\kappa}\left[\kappa^{-(M-K+1)}\right]$ is finite we achieve full diversity. In the following we will analyse this term. Using variable κ definition we get

$$
\begin{aligned}
&\mathbb{E}_{\kappa}\left[\kappa^{-(M-K+1)}\right] \\
&= \mathbb{E}_{\mathcal{N}_L[\mathbf{H}_{00}]}\left[\left(\frac{1}{T} tr(\mathcal{N}_L[\mathbf{H}_{00}]^H \mathbf{D}_{\mathbf{V}} \mathcal{N}_L[\mathbf{H}_{00}])\right)^{-(M-K+1)}\right]
\end{aligned}
\tag{51}
$$

Assume, for now, that $L = 2$ entries of matrix $\mathbf{D}_{\mathbf{V}}$ are equal to d and the others are zero, i.e. $\mathbf{D}_{\mathbf{V}} = diag(d, d, 0, \ldots, 0)$. d denotes the smallest singular value of the correlation matrix ($\mathbf{V}\mathbf{V}^H$). Let $\mathbf{D}_1 = diag(0, \ldots, 0, 1, \ldots, 1)$, $tr(\mathbf{D}_1) = M - K$, $\mathbf{D}_2 = diag(1, 1, 0, \ldots, 0)$ and $tr(\mathbf{D}_2) = 2$, then

$$
\begin{aligned}
&\mathbb{E}_{\kappa}\left[\kappa^{-(M-K+1)}\right] \\
&= \mathbb{E}_{\mathbf{U}_{00}}\left[\left(\frac{d}{T} tr(\mathbf{D}_1 \mathbf{U}_{00}^H \mathbf{D}_2 \mathbf{U}_{00})\right)^{-(M-K+1)}\right] \\
&\leq \mathbb{E}_{x,y}\left[\left(\frac{d}{T}(1 - (1-x)(1-y))\right)^{-(M-K+1)}\right]
\end{aligned}
\tag{52}
$$

The inequality follows from Lemma 3 equation (18). From (19), i.e. the joint probability density function (pdf) of x and y

$$\mathbb{E}_\kappa \left[\kappa^{-(M-K+1)} \right]$$
$$\leq \frac{M-1}{K-1} \left(\frac{(M-2)!}{(M-K-1)!(K-2)!} \right)^2 \frac{(M-K)^{-2}}{(M-K-1)} \tag{53}$$

Therefore, for $M-K > 1$ (for $M-K = 1$ bound (53) is infinite) full diversity is achievable

$$\overline{BER}(P_0, \mathcal{T}) \dot{\leq} \mathbb{E}_\kappa \left[\left(\kappa \frac{P_0}{\sigma^2} \right)^{-(M-K+1)} \right] \doteq \left(\frac{P_0}{\sigma^2} \right)^{-(M-K+1)} \tag{54}$$

For $K = M-1$ the behavior of the BER curve is different and the previous analysis is not able to capture it. Let us consider then the case $K = M-1$. For $K = M-1$ the pdf of the minimum eigenvalue of a $M \times (M-1)$ MIMO channel ($\lambda_{M-1}[\mathbf{H}_{00}]$) is [28], [36], [37]

$$f_{\Lambda_{M-1}}(\lambda_{M-1}) = \beta^2 \lambda_{M-1}[\mathbf{H}_{00}] e^{-\beta \lambda_{M-1}[\mathbf{H}_{00}]} \tag{55}$$

where β is positive and real. Using the exponential upper bound of the Q-function the variable κ definition and evaluating the expectation operator of equation (49) over $\lambda_K[\mathbf{H}_{00}]$ we get

$$\overline{BER}(P_0, \mathcal{T})$$
$$\leq \mathbb{E}_\kappa \left[\frac{1}{2} \left(1 + \xi \kappa \frac{\beta^{-1}}{2} \frac{P_0}{\sigma^2} \right)^{-2} \right]$$
$$= \mathbb{E}_{\mathbf{u}_{00}^0} \left[\frac{1}{2} \left(1 + \frac{\xi \beta^{-1}}{2T} \frac{P_0}{\sigma^2} \mathbf{u}_{00}^{0}{}^H \mathbf{D_V} \mathbf{u}_{00}^0 \right)^{-2} \right] \tag{56}$$

where $\mathbf{u}_{00}^0 = \mathcal{N}_L[\mathbf{H}_{00}] \in \mathbb{C}^{M \times 1}$. As $\mathbf{u}_{00}^{0}{}^H \mathbf{u}_{00}^0 = 1$ we have

$$\overline{BER}(P_0, \mathcal{T}) \leq \frac{1}{2} \mathbb{E}_{\mathbf{u}_{00}^0} \left[\left(\mathbf{u}_{00}{}^H \mathbf{D} \mathbf{u}_{00} \right)^{-2} \right] \tag{57}$$

where $\mathbf{D} = \mathbf{I} + \frac{\xi \beta^{-1}}{2T} \frac{P_0}{\sigma^2} \mathbf{D_V}$. As before, assume that L entries of matrix $\mathbf{D_V}$ are equal to d and the others are zero, i.e. $\mathbf{D} = diag(a, \ldots, a, 0, \ldots, 0)$, where $a = 1 + \frac{d\xi\beta^{-1}}{2T} \frac{P_0}{\sigma^2}$. As previously stated, $\mathbf{u}_{00}^0$ is uniformly distributed over the set of unitary vectors, with dimension M [37], [36]. Therefore $\mathbf{u}_{00}^0 \sim \mathbf{h}/\sqrt{\mathbf{h}^H \mathbf{h}}$, where $\mathbf{h} \in \mathbb{C}^{M \times 1}$ is i.i.d. complex Gaussian distributed with unitary variance. Set $x = \sum_{m=1}^{L} |\mathbf{h}(m)|^2$ and $y = \sum_{m=L+1}^{M} |\mathbf{h}(m)|^2$, then bound (57) simplifies to

$$\overline{BER}(P_0, \mathcal{T}) \leq \frac{1}{2} \mathbb{E}_{x,y} \left[\left(\frac{x+y}{ax+y} \right)^2 \right] \tag{58}$$

The random variables x and y are the sum of L and $M-L$ exponentially distributed random variables, respectively, then their distribution is as follows

$$f_X(x) = \frac{x^{L-1} e^{-x}}{(L-1)!}, \quad f_Y(y) = \frac{y^{M-L-1} e^{-y}}{(M-L-1)!} \tag{59}$$

Consider the case where $L = 2$. As x is always equal to or higher than zero then $(x+y)^{M-3} \geq y^{M-3}$. For this reason

the expectation operator can be upper bounded by

$$\mathbb{E}_{x,y} \left[\left(\frac{x+y}{ax+y} \right)^2 \right]$$
$$\leq \int_0^{+\infty} \int_0^{+\infty} \frac{(x+y)^{M-1}}{(ax+y)^2} \frac{xe^{-x}e^{-y}}{(M-3)!} dxdy$$
$$= (2 + (M-3)M) \frac{1 - a + a\log(a)}{a(a-1)^2}$$
$$\leq (2 + (M-3)M) \frac{\log(a)}{(a-1)^2} \tag{60}$$

The last inequality follows from the definition of a and d. Hence, for the case $L = 2$ the BER can be upper bounded by

$$\overline{BER}(P_0, \mathcal{T})$$
$$\leq \frac{2 + (M-3)M}{2} \log \left(1 + \frac{d\xi\beta^{-1}}{2T} \frac{P_0}{\sigma^2} \right) \left(\frac{d\xi\beta^{-1}}{2T} \frac{P_0}{\sigma^2} \right)^{-2}$$
$$\doteq \log \left(\frac{P_0}{\sigma^2} \right) \left(\frac{\sigma^2}{P_0} \right)^2 \tag{61}$$

Now, let us consider the case where $L > 2$. From the definition of y we have that $y \geq 0$ then $ax + y \geq ax$. As a result the expectation operator for $L > 2$ is upper bounded by

$$\mathbb{E}_{x,y} \left[\left(\frac{x+y}{ax+y} \right)^2 \right] \leq \mathbb{E}_{x,y} \left[\left(\frac{x+y}{ax} \right)^2 \right]$$
$$= \frac{2 + (M-3)M}{2 + (L-3)L} a^{-2} \tag{62}$$

Thus, for the case where $L > 2$ the BER is upper bounded as follows

$$\overline{BER}(P_0, \mathcal{T}) \leq \frac{1}{2} \frac{2 + (M-3)M}{2 + (L-3)L} \left(1 + \frac{d\xi\beta^{-1}}{2T} \frac{P_0}{\sigma^2} \right)^{-2}$$
$$\doteq \left(\frac{\sigma^2}{P_0} \right)^2 \tag{63}$$

APPENDIX F
PROOF LEMMA 4

Consider a Rayleigh distributed channel $\mathbf{h} \in \mathbb{C}^{M \times 1}$ and a diagonal matrix $\mathbf{D} \in \mathbb{C}^{M \times M}$. Define the following parameter

$$c = \mathbb{E}_\mathbf{h}[e^{-\frac{P_0}{\sigma^2} \mathbf{h}^H \mathbf{D} \mathbf{h}}] = \mathbb{E}_\mathbf{h} \left[e^{-\frac{P_0}{\sigma^2} \sum_{m=1}^{M} \mathbf{D}(m,m) x_m} \right] \tag{64}$$

where $x_m = |\mathbf{h}(m)|^2$ is exponentially distributed. By evaluating the expectation operator

$$c = \prod_{m=1}^{M} \left(1 + \frac{P_0}{\sigma^2} \mathbf{D}(m,m) \right)^{-1}$$
$$= \left| \mathbf{I}_M + \frac{P_0}{\sigma^2} \mathbf{D} \right|^{-1} \doteq \left(\frac{\sigma^2}{P_0} \right)^M |\mathbf{D}|^{-1} \tag{65}$$

As in the proof of theorem 2 set $\mathbf{u}_1 = \mathbf{h}/\sqrt{\mathbf{h}^H \mathbf{h}}$. Set also $\mu = \mathbf{h}^H \mathbf{h}$. As μ is the sum of M exponential distributed

random variables, its pdf is $f_U(\mu) = \frac{\mu^{M-1}}{(M-1)!}e^{-\mu}$. Therefore, the previous expectation can also be evaluated as follows

$$
\begin{aligned}
c &= \mathbb{E}_{\mathbf{h}}\left[e^{-\frac{P_0}{\sigma^2}\mathbf{h}^H\mathbf{D}\mathbf{h}}\right] = \mathbb{E}_{\mu,\mathbf{u}_1}\left[e^{-\mu\frac{P_0}{\sigma^2}\mathbf{u}_1^H\mathbf{D}\mathbf{u}_1}\right]\\
&= \mathbb{E}_{\mathbf{u}_1}\left[\int_0^{+\infty}e^{-\mu\frac{P_0}{\sigma^2}\mathbf{u}_1^H\mathbf{D}\mathbf{u}_1}f_U(\mu)d\mu\right]\\
&= \mathbb{E}_{\mathbf{u}_1}\left[\left(1+\frac{P_0}{\sigma^2}\mathbf{u}_1^H\mathbf{D}\mathbf{u}_1\right)^{-M}\right]\\
&\doteq \left(\frac{\sigma^2}{P_0}\right)^M\mathbb{E}_{\mathbf{u}_1}\left[(\mathbf{u}_1^H\mathbf{D}\mathbf{u}_1)^{-M}\right]
\end{aligned}
\tag{66}
$$

The third equality follows from the fact that μ and $\mathbf{u}_1$ are independent random variables [36], [37]. From (65) and (66) we verify that

$$
\mathbb{E}_{\mathbf{u}_1}\left[(\mathbf{u}_1^H\mathbf{D}\mathbf{u}_1)^{-M}\right]\doteq|\mathbf{D}|^{-1}
\tag{67}
$$

As $f(x)=x^{\alpha}$ is concave for all $\alpha < 1$ then by applying Jensen inequality to the expectation operation of equation (50) we get for $\alpha = (M-K+1)/M$

$$
\overline{BER}(P_0,\mathcal{T})\dot{\leq}\mathbb{E}_{\kappa}\left[\kappa^{-M}\right]^{\frac{M-K+1}{M}}\left(\frac{P_0}{\sigma^2}\right)^{-(M-K+1)}
\tag{68}
$$

Let $\mathcal{N}_L[\mathbf{H}_{00}] = [\mathbf{b}_1,\ldots,\mathbf{b}_m,\ldots,\mathbf{b}_{M-K}]$, where $\mathbf{b}_m \in \mathbb{C}^{M\times 1}$ denotes column m of matrix $\mathcal{N}_L[\mathbf{H}_{00}]$ and consider that $L \geq M$, then from the definition of variable κ we obtain

$$
\kappa = \frac{1}{T}tr(\mathcal{N}_L[\mathbf{H}_{00}]^H\mathbf{D_V}\mathcal{N}_L[\mathbf{H}_{00}]) \leq \frac{1}{T}\mathbf{b}_1^H\mathbf{D_V}\mathbf{b}_1
\tag{69}
$$

Therefore, from (67) the BER is upper bounded by

$$
\begin{aligned}
&\overline{BER}(P_0,\mathcal{T})\\
&\dot{\leq}\mathbb{E}_{\kappa}\left[(\mathbf{b}_1^H\mathbf{D_V}\mathbf{b}_1)^{-M}\right]^{\frac{M-K+1}{M}}\left(\frac{P_0}{\sigma^2}\right)^{-(M-K+1)}\\
&= |\mathbf{V}\mathbf{V}^H|^{-\frac{M-K+1}{M}}\left(\frac{P_0}{\sigma^2}\right)^{-(M-K+1)}
\end{aligned}
\tag{70}
$$

The equality follows from (67) and $|\mathbf{D_V}| = |\mathbf{V}\mathbf{V}^H|$, since $\mathbf{u}_1 \sim \mathbf{b}_1$ by definition.

APPENDIX G
PROOF THEOREM 3

As the objective function of both problems ((24) and (25)) is the same and the constraint from problem (25) includes all points in the feasibility region, defined by the constraint set of problem (24), then the optimum value of optimization problem (25) constitutes an upper bound to the optimum value of problem (24).

Consider the relaxed optimization problem (25) and that the SVD of $\mathbf{V}$ is as in definition 6. Thus the SVD of $\mathbf{V}\mathbf{V}^H$ is given by $\mathbf{U_V}\mathbf{D_V}\mathbf{U_V}^H$, where $\mathbf{D_V} = \mathbf{\Lambda_V}^H\mathbf{\Lambda_V} = \mathbf{\Lambda_V^1}^H\mathbf{\Lambda_V^1}$. Using the Sylvester's determinant theorem and the fact that the constraint $\sum_{t=1}^T\mathbf{V}^H(t)\mathbf{V}(t) = T$ is equivalent to $tr(\mathbf{V}^H\mathbf{V}) = T$ problem (25) can be rewritten as

$$
\max_{\mathbf{V}}\prod_{m=1}^M\mathbf{D_V}(m,m),\quad s.t.\quad\sum_{m=1}^M\mathbf{D_V}(m,m) = T
\tag{71}
$$

Therefore the left and right singular vectors of $\mathbf{V}$ have no impact on the optimum value of the relaxed optimization problem. Applying the Lagrangian multiplier method the optimum solution for (71) is $\mathbf{D_V}(p,p) = T/M$ and the respective optimum value is $(T/M)^M$. Consequently, the optimum value for $\mathbf{\Lambda}_V$ is $\left[\sqrt{T/M}\mathbf{I}_M\ \mathbf{0}\right]$.

Now, consider the optimization problem (24). The constraints $\mathbf{V}(t)^H\mathbf{V}(t) = 1$, $t \in \{1,\ldots,T\}$ are equivalent to the constraint that $\mathbf{V}^H\mathbf{V}$ must be a unit diagonal matrix. Theorem 2.1 from [30] states a necessary and sufficient condition for the existence of a unit diagonal matrix with a given set of eigenvalues. This condition is respected for the singular values $\mathbf{D_V}(p,p) = T/M$, $p \in \{1,\ldots,M\}$. As a consequence we have proved the existence of a matrix $\mathbf{V}$ with eigenvalues equal to the ones provided by the optimum of the relaxed problem and that in addition respects the constraints of the original problem (24). Since both optimization problems share the same cost function and the optimum of the relaxed problem is an upper bound to the original problem then the optimum of the last may be obtained from the first.

To generate matrix $\mathbf{V}^H\mathbf{V}$ with unit diagonal and fixed eigenvalues we can use the algorithm 3.1 described in [30].

REFERENCES

[1] J. Hoadley and P. Maveddat, "Enabling Small Cell Deployment with HetNet," *IEEE Wireless Commun. Mag.*, vol. 19, no. 2, pp. 4–5, 2012.

[2] T. Nakamura, S. Nagata, A. Benjebbour, Y. Kishiyama, T. Hai, S. Xiaodong, Y. Ning, and L. Nan, "Trends in Small Cell Enhancements in LTE Advanced," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 98–105, 2013.

[3] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking Spectrum Gridlock with Cognitive Radios: An Information Theoretic Perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, 2009.

[4] S. Perlaza, N. Fawaz, S. Lasaulce, and M. Debbah, "From Spectrum Pooling to Space Pooling: Opportunistic Interference Alignment in MIMO Cognitive Networks," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3728–3741, 2010.

[5] M. Matinmikko, M. Mustonen, M. Hoyhtya, T. Rauma, H. Sarvanko, and A. Mammela, "Distributed and Directional Spectrum Occupancy Measurements in the 2.4 Ghz ISM band," Proc. *IEEE Wireless Communication Systems (ISWCS)*, 2010, pp. 676–980.

[6] A. Attar, V. Krishnamurthy, and O. Gharehshiran, "Interference Management using Cognitive Base-Stations for UMTS LTE," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 152–159, 2011.

[7] J. Xie, Z. Fu, and H. Xian, "Spectrum Sensing Based on Estimation of Direction of Arrival," Proc. *IEEE Computational Problem-Solving (ICCP)*, 2010, pp. 39–42.

[8] K. Cumanan, L. Musavian, S. Lambotharan, and A. Gershman, "SINR Balancing Technique for Downlink Beamforming in Cognitive Radio Networks," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 133–136, 2010.

[9] Y. J. Zhang and A. M.-C. So, "Optimal Spectrum Sharing in MIMO Cognitive Radio Networks Via Semidefinite Programming," *IEEE J Sel Areas Commun*, vol.29, no.2, pp.362,373, February 2011

[10] W. Shin, W. Noh, K. Jang, and H.-H. Choi, "Hierarchical Interference Alignment for Downlink Heterogeneous Networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4549–4559, 2012.

[11] M. Maso, M. Debbah, and L. Vangelista, "A Distributed Approach to Interference Alignment in OFDM-based Two-Tiered Networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1935–1949, 2013.

[12] S. Sharma, S. Chatzinotas, and B. Ottersten, "Interference Alignment for Spectral Coexistence of Heterogeneous Networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, p. 46, 2013.

[13] V. Cadambe and S. Jafar, "Interference Alignment and Degrees of Freedom of the k -user Interference Channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, 2008.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2014.2322855, IEEE Transactions on Wireless Communications

13

[14] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X Channels: Interference Alignment, Decomposition, and Performance Analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, 2008.

[15] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the Degrees of Freedom Achievable through Interference Alignment in a MIMO Interference Channel," *Signal Processing, IEEE Transactions on*, vol. 60, no. 2, pp. 812–821, Feb 2012.

[16] K. Gomadam, V. Cadambe, and S. Jafar, "A Distributed Numerical Approach to Interference Alignment and Applications to Wireless Interference Networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, 2011.

[17] S. Peters and R. Heath, "Interference Alignment via Alternating Minimization," Proc. *IEEE Acoustics, Speech and Signal Processing (ICASSP)*, 2009. pp. 2445–2448.

[18] D. Kim and M. Torlak, "Optimization of Interference Alignment Beamforming Vectors," *IEEE J Sel Areas Commun*, vol. 28, no. 9, pp. 1425–1434, 2010.

[19] R. Krishnamachari and M. Varanasi, "Interference Alignment Under Limited Feedback for MIMO Interference Channels," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3908–3917, 2013.

[20] M. Rezaee and M. Guillaud, "Interference Alignment with Quantized Grassmannian Feedback in the K-user MIMO Interference Channel," *arXiv preprint arXiv:1207.6902 [cs.IT] [Online]. Available: http://arxiv.org/ abs/1207.6902*, 2012.

[21] O. E. Ayach and R. W. Heath, "Interference Alignment with Analog Channel State Feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, 2012.

[22] H.-H. Lee and Y.-C. Ko, "Interference Alignment with Random Vector Quantization for MIMO Interference Channels," Proc. *IEEE Vehicular Technology Conference (VTC Fall)*, 2012, pp. 1–5.

[23] O. El Ayach and R. Heath, "Grassmannian Differential Limited Feedback for Interference Alignment," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6481–6494, 2012.

[24] L. Zheng and D. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.

[25] P. Cordier, P. Houze, S. Jemaa, O. Simon, D. Bourse, D. Grandblaise, K. Moessner, J. Luo, C. Kloeck, K. Tsagkaris *et al.*, "E2R Cognitive Pilot Channel Concept," Proc. *Mobile and Wireless Communications Summit (IST)*, 2006, pp. 1–4.

[26] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2001.

[27] S. Benedetto and E. Biglieri, *Digital Transmission Theory: With Wireless Applications*. Springer, 1999.

[28] E. Sengul, E. Akay, and E. Ayanoglu, "Diversity Analysis of Single and Multiple Beamforming," *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 990–993, 2006.

[29] A. Edelman and N. R. Rao, "Random matrix Theory," *Acta Numerica*, vol. 14, no. 1, pp. 233–297, 2005.

[30] P. I. Davies and N. J. Higham, "Numerically Stable Generation of Correlation Matrices and their Factors," *BIT Numerical Mathematics*, vol. 40, no. 4, pp. 640–651, 2000.

[31] C. K. Au-Yeung and D. Love, "On the Performance of Random Vector Quantization Limited Feedback Beamforming in a MISO System," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, 2007.

[32] D. J. Love, R. W. Heath Jr, and T. Strohmer, "Grassmannian Beamforming for Multiple-Input Multiple-Output Wireless Systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.

[33] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 2012.

[34] Y. Jiang, M. Varanasi, and J. Li, "Performance Analysis of ZF and MMSE Equalizers for MIMO Systems: An in-depth Study of the High SNR Regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2008–2026, 2011.

[35] L. Garcia-Ordonez, A. Pages-Zamora, and J. Fonollosa, "Diversity and Multiplexing tradeoff of Spatial Multiplexing MIMO Systems with CSI," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2959–2975, 2008.

[36] T. Ratnarajah, R. Vaillancourt, and M. Alvo, "Eigenvalues and Condition Numbers of Complex Random Matrices," *SIAM J. Matrix Anal. Applic.*, vol. 26, no. 2, pp. 441–456, 2004.

[37] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices," *Ph.D. Dissertation, Mass. Inst. Technol.*, Cambridge, MA, 1989.